

# ***Draft Systematic Review***

---

**Number XX**

## **Diagnosis of Gout**

**Prepared for:**

Agency for Healthcare Research and Quality  
U.S. Department of Health and Human Services  
540 Gaither Road  
Rockville, MD 20850  
[www.ahrq.gov](http://www.ahrq.gov)

**Contract No.**

**Prepared by:**

This information is distributed solely for the purposes of predissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.

**Investigators:**

**AHRQ Publication No. xx-EHCxxx  
<Month Year>**

This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This report may periodically be assessed for the urgency to update. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov). Search on the title of the report.

This document is in the public domain and may be used and reprinted without special permission. Citation of the source is appreciated.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact [EffectiveHealthCare@ahrq.hhs.gov](mailto:EffectiveHealthCare@ahrq.hhs.gov).

None of the investigators has any affiliations or financial involvement related to the material presented in this report.
---

**Suggested citation:** To Be Added for Final Version

## Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see [www.effectivehealthcare.ahrq.gov/reference/purpose.cfm](http://www.effectivehealthcare.ahrq.gov/reference/purpose.cfm)

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site ([www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov)) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to [epc@ahrq.hhs.gov](mailto:epc@ahrq.hhs.gov).

Richard G. Kronick, Ph.D.  
Director  
Agency for Healthcare Research and Quality

David Meyers, M.D.  
Acting Director, Center for Evidence and  
Practice Improvement  
Agency for Healthcare Research and Quality

Stephanie Chang M.D., M.P.H.  
Director, EPC Program  
Center for Evidence and Practice  
Improvement  
Agency for Healthcare Research and Quality

Aysegul Gozu, M.D., M.P.H.  
Task Order Officer  
Center for Evidence and Practice  
Improvement  
Agency for Healthcare Research and Quality

## **Acknowledgments**

To Be Added for Final Version

## **Key Informants**

To Be Added for Final Version

## **Technical Expert Panel**

To Be Added for Final Version

## **Peer Reviewers**

To Be Added for Final Version

# Diagnosis of Gout

## Structured Abstract

**Objectives.** The aim of this review is to assess the evidence for the accuracy and safety of tests for diagnosis of gout, including algorithms combining clinical signs and symptoms, dual emission computerized tomography (DECT), ultrasound (US), and plain x-ray, and factors that affect accuracy, with particular emphasis on tests that can be conducted in primary and acute care settings.

**Data Sources.** We searched PubMed (from 1946 to the present), EMBASE (from 1972), the Cochrane Collection (from 1945), and the Web of Science (from 1949) for published studies. We also searched Clinicaltrials.gov and the Web of Science for unpublished data. Manufacturers of imaging equipment and test kits were contacted for unpublished data on their use for gout diagnosis.

**Review Methods.** Published observational studies that reported on the accuracy and safety of diagnostic tests in comparison to an accepted reference standard in gout suspects and individuals without gout, studies that reported on factors affecting diagnostic accuracy, studies that reported on outcomes of misdiagnosis of gout, and unpublished data identified through grey literature searches or provided by manufacturers for accuracy or safety outcomes were included, as were recent systematic reviews that reported on outcomes of interest and that met the inclusion criteria. A standardized protocol with predefined criteria was used to extract details on study design, interventions, outcomes, and study quality and to assess the strength of evidence for each conclusion.

**Results.** Several algorithms comprising clinical signs and symptoms have been tested for diagnostic accuracy against the presence of monosodium urate (MSU) crystals joint aspirate, or another set of criteria; most studies are conducted with small groups of patients in secondary care settings. One particular algorithm, the Diagnostic Rule, which is the only one developed and validated with primary care physicians and patients, has an area under the curve (AUC) of 0.85 and is simple to administer. Three studies of DECT that enrolled patients without a previous gout diagnosis revealed sensitivities ranging from 85% to 100% and specificities ranging from 83% to 92% in diagnosing gout. Four studies of ultrasound (US) that enrolled patients without a previous diagnosis showed sensitivities ranging from 37% to 100% and specificities ranging from 68% to 97%, depending on the signs assessed. No studies examined factors that affected the accuracy of clinical algorithms, DECT, or US for the diagnosis of gout. The accuracy of MSU analysis is affected by factors including practitioner skill, presentation time of the patient and sample handling. No studies reported adverse events associated with techniques used to diagnose gout; however in one study, misdiagnosis of gout was shown to result in unnecessary surgery and delay in appropriate treatment.

**Conclusions.** Promising diagnostic algorithms such as the Diagnostic Rule need to be validated in primary care settings. Both DECT and US show good sensitivity and specificity for gout

diagnosis in high risk patients. An algorithm with high diagnostic accuracy can ideally form part of a decision tree that combines clinical signs and symptoms with more invasive tests or imaging for clinically ambiguous cases.

# Contents

<b>Executive Summary .....</b>	<b>ES-1</b>
<b>Introduction.....</b>	<b>1</b>
Background .....	1
Scope and Key Questions .....	2
Key Question 1. ....	3
Key Question 2. ....	4
Organization of this report .....	4
<b>Methods.....</b>	<b>5</b>
Criteria for Inclusion/Exclusion of Studies in the Review .....	5
Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions.....	6
Assessment of Methodological Quality of Individual Studies .....	7
Grading the Strength of the Body of Evidence for Each Key Question .....	8
Peer Review and Public Commentary .....	9
<b>Results .....</b>	<b>10</b>
Introduction.....	10
Results of Literature Searches .....	10
Key Question 1a-c: .....	12
Key Points.....	12
Description of included studies.....	12
Detailed Synthesis.....	13
KQ1d. Factors Affecting the Accuracy of Synovial Fluid Aspiration and Crystal Analysis ...	40
Key Points.....	40
Description of Included Studies.....	40
Detailed Synthesis.....	40
Key Question 2: What are the adverse effects associated with each diagnostic test (including pain, infection at the aspiration site, radiation exposure) or harms (related to false positives, false negatives, indeterminate results) associated with tests used to diagnose gout? .....	42
Key Points.....	42
Description of included studies.....	42
Detailed Synthesis.....	42
<b>Discussion.....</b>	<b>44</b>
Key Findings and Strength of Evidence .....	44
Findings in Relationship to What is Already Known .....	45
Applicability .....	48
Implications for Clinical and Policy Decisionmaking .....	48
Limitations of the Comparative Effectiveness Review Process .....	48
Limitations of the Evidence Base .....	49
Research Gaps.....	50
Conclusions.....	50
<b>References.....</b>	<b>52</b>
<b>Abbreviations / Acronyms.....</b>	<b>56</b>

## Tables

Table A. Comparison of Components among Clinical Algorithms for Diagnosis of Gout....	ES-12
Table B. Summary of Findings and Strength of Evidence .....	ES-16
Table 1. Summary of Included Studies of Algorithms Comprising Clinical Signs and Symptoms Used to Diagnose Gout .....	17
Table 2. Comparison of Components Among Clinical Algorithms for Diagnosis of Gout .....	22
Table 3. Studies Assessing the Accuracy of DECT to Diagnose Gout .....	26
Table 4. Summary of Studies Reporting on the Use of Ultrasound for Diagnosis of Gout.....	32
Table 5. Risk of bias assessment by QUADAS-2.....	38
Table 6. Summary of Findings and Strength of Evidence .....	50

## Figures

Figure A. Analytic Framework for the Diagnosis of Gout .....	ES-3
Figure B. Literature flow diagram .....	ES-9
Figure 1. Analytic Framework .....	3
Figure 2. Literature flow diagram .....	11

## Appendices

Appendix A: Search Strategy
Appendix B: List of Excluded Studies
Appendix C. Evidence Table
Appendix D. Data Abstraction Tools





# Executive Summary

## Background

### Condition

Gout is a form of inflammatory arthritis characterized by acute intermittent episodes of synovitis presenting with joint swelling and pain (referred to as acute gouty arthritis) that may progress to a chronic intermittent condition, which may progress further to development of tophi (solid deposits of monosodium urate [MSU] crystals in joints, cartilage, and bones), a condition called chronic tophaceous gout.

Based on data from the 2007-2008 National Health and Nutrition Examination Survey (NHANES), the prevalence of gout among adults in the United States has been estimated to be 3.9 percent (8.3 million individuals), ranging from 2.0 percent in women to 5.9 percent in men.<sup>1</sup> Comparing the most recent figures for the prevalence of gout to those of previous cycles of NHANES shows that the prevalence of gout appears to be increasing. The rise in the prevalence of gout has paralleled the increase in prevalence of conditions associated with hyperuricemia, including obesity, hypertension, hypertriglyceridemia, hypercholesterolemia, type 2 diabetes and metabolic syndrome, chronic kidney disease, and renal insufficiency. Certain medications also may increase the risk for developing gout (e.g., thiazide diuretics).

A 2013 study estimated the annual costs of gout to be \$933 million (in 2008 figures), with the annual ambulatory care costs associated with gout potentially reaching \$1 billion. Some 32 percent of the costs were attributed to gouty arthritis attacks, and drug expenditures accounted for 61 percent of the total costs.<sup>1</sup>

**Etiology of Gout.** The driving force behind acute episodes of gout is hyperuricemia (defined as a serum uric acid (strictly, urate, sUA) concentration greater than 6.8 mg per deciliter [dl] in men and greater than 6.0 in women). Hyperuricemia can be the result of either inadequate renal excretion of UA or, less commonly, UA overproduction (UA is a breakdown product of dietary or endogenous purines, which are among the building blocks of nucleic acids); and is associated with the formation and deposition of the UA crystals, which preferentially dissolve, in joints, tendons, and bursa spaces. Despite the prevalence of hyperuricemia, for reasons that remain unclear, only a small proportion of individuals with hyperuricemia go on to develop gout; in the rest, hyperuricemia remains asymptomatic.<sup>2</sup> The prevalence of hyperuricemia ranges from 21.2 percent in men to 21.6 percent in women, four- to ten-fold higher than the prevalence of gout.<sup>3</sup>

The causes of gout are multifactorial, including a combination of genetic, hormonal, metabolic, pharmacologic, renal disease, and dietary factors. Family history, advancing age, male sex, or, in women, early menopause have been associated with a higher risk of gout and/or gout flares.<sup>4</sup> Dietary risk factors for gout appear to include alcohol consumption, as well as consumption of meat, seafood, sugar sweetened soft drinks, and foods high in fructose, whereas dairy foods and coffee have been associated with a lower risk of incident gout and in some cases a lower rate of gout flares. However, the role of diet in the etiology and treatment of gout is a topic of considerable research and will be reviewed in a separate systematic review.

## Diagnostic Strategies

Some research has supported the need for laboratory assessment of joint/synovial fluid MSU in the setting of an acute inflammatory arthritis for a definitive diagnosis of gout, a test generally conducted by a rheumatologist to confirm a conditional/presumptive diagnosis.<sup>5</sup> However, the majority of individuals with gout are initially seen, diagnosed, and treated in primary and urgent care settings. Thus primary care physicians (PCPs) and emergency medicine physicians are the most likely practitioners to see patients with early-stage gout. However, use of the gold standard synovial fluid analysis for diagnosis of gout is difficult and seldom performed in the primary or emergent care setting.<sup>6,7</sup> In fact, evidence from a 2011 survey of rheumatologists suggests that SF analysis is underused in the rheumatology setting as well.<sup>8</sup> Instead, PCPs and emergency medicine physicians may tend to rely on an algorithm comprising some combination of clinical signs and symptoms to diagnose an acute episode of gout. Attempts to standardize and validate such diagnostic algorithms date back to the 1970s.<sup>9</sup> A question of interest is whether any combination of clinical signs and symptoms and laboratory tests accessible in the primary or acute care setting will have good predictive value compared with tests such as joint aspiration and synovial fluid analysis. Plain radiographs have also been used for gout diagnosis, and newer techniques, including ultrasound (US) and dual-energy computed tomography (DECT) are just beginning to be used to diagnose gout in some settings.<sup>10</sup> Another question of interest is how these newer methods compare with joint aspiration and synovial fluid analysis in their predictive value for the diagnosis of gout.

Therefore, we have undertaken a systematic review of studies examining the accuracy of tests used to diagnose gout, including combinations of physical findings, serum UA, US, plain radiography, and DECT, compared with synovial fluid UA. The results of this review should help inform clinical decision-making for patients and providers and improve the quality of care for patients with gout in the primary and acute care setting.

## Scope and Key Questions

### Scope of the Review

The purpose of this review is to assess the evidence on the comparative validity and safety of tests used for the diagnosis of gout, including clinical signs and symptoms (individually and in combination as an algorithm), DECT, US, and other visualization methods, compared with the gold standard of aspiration of synovial fluid from involved joints and analysis of monosodium urate (MSU) crystals using polarized light microscopy. The review also assesses the evidence that practitioner type and other factors may affect the outcomes of MSU analysis. AHRQ assigned this report to the Southern CA Evidence-based Practice Center (HHS290201200006I). A protocol for the review was posted on the AHRQ website on July 17, 2014 at: <http://effectivehealthcare.ahrq.gov/ehc/products/564/1937/gout-protocol-140716.pdf>. The protocol was approved by the AHRQ Center for Evidence and Practice Improvement.

## Key Questions

Figure A shows an analytic framework to illustrate the population, interventions, outcomes, and adverse effects that will guide the literature search and synthesis for this project.

**Figure A. Analytic Framework for the Diagnosis of Gout**

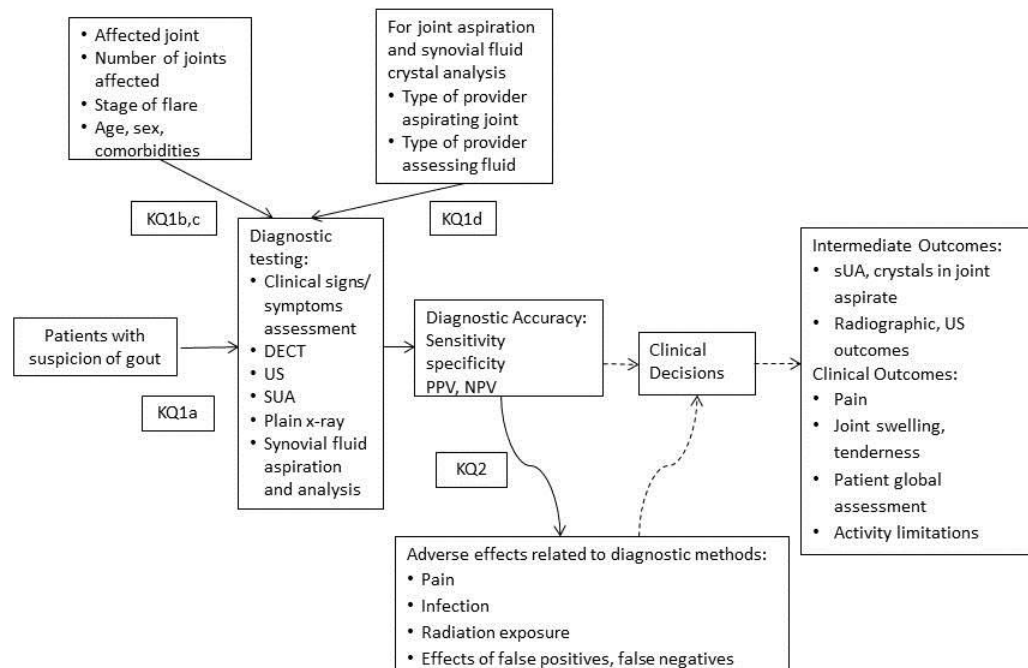


Figure notes: KQ=Key Question; NPV=Negative predictive value; PPV=Positive predictive value; sUA=Serum uric acid; US=Ultrasound

## Key Question 1

- What is the accuracy of clinical signs and symptoms and other diagnostic tests (such as serum uric acid, ultrasound, CT scan, DECT, and plain x-ray), alone or in combination, compared to synovial fluid analysis in the diagnosis of acute gouty arthritis, and how does the accuracy affect clinical decision making, clinical outcomes and complications, and patient centered outcomes?
- How does the diagnostic accuracy of clinical signs and symptoms and other tests vary by affected joint site and number of joints?
- Does the accuracy of diagnostic tests for gout vary by duration of symptoms (i.e., time from the beginning of a flare)?
- Does the accuracy of synovial fluid aspiration and crystal analysis differ by i) the type of practitioner who is performing the aspiration and ii) the type of practitioner who is performing the crystal analysis?

## Key Question 2.

What are the adverse effects associated with each diagnostic test (including pain, infection at the aspiration site, radiation exposure) or harms (related to false positives, false negatives, indeterminate results) associated with tests used to diagnose gout?

## Methods

### Criteria for Inclusion/Exclusion of Studies in the Review

This report is based on a systematic search for prospective or cross-sectional studies that compared the sensitivity/specificity of tests used to diagnose gout against the gold standard test of joint aspiration and synovial fluid assessment for MSU crystals in populations of adults 18 years of age or older, suspected of having gout (see Table A, below). Tests of interest included clinical examination for physical signs, symptoms, and history; serum uric acid, ultrasonography; DECT; and plain radiography. The comparator of primary interest was synovial fluid analysis of MSU crystals using polarized light microscopy. Studies were also included if some or all of the index test comparisons were to the American College of Rheumatology criteria for gout diagnosis as a reference standard (comparator). Studies were excluded if enrolled participants had already been definitively diagnosed with gout, if the patients had chronic gout, or if the comparator was individual physician opinion or was not identified.

Outcomes of interest were the accuracy of the test results (the sensitivity and specificity or the positive and negative predictive value of the test in question), intermediate outcomes such as lab and radiographic test results, clinical decision making, short term clinical (patient-centered) outcomes such as pain and joint swelling, and any adverse events (including adverse patient experiences such as pain or infection at the aspiration site, effects of radiation exposure, and the results of a false positive or negative) associated with the test. Studies were also accepted if they examined factors that potentially affected the validity of tests (including but not limited to joints involved, duration of symptoms, sex, or types of practitioners performing or evaluating the tests).

Prospective cohort, cross-sectional, and case-control (if needed) studies that compared the sensitivity/specificity or area under the curve between a proposed diagnostic test and the gold standard were included to address KQ1a. Studies of similar design that compared outcomes based on the joint(s) involved (KQ1b), the lag between the onset of symptoms and diagnosis (commencement of treatment) (KQ1c) or the type of practitioner who conducted or analyzed the test (or other similar factors affecting test results) were also included. Prospective cohort, case control, and case series of any size, as well as case reports of rare adverse events were included if they addressed adverse events or other negative outcomes in individuals undergoing testing (KQ2).

### PICOTS

#### Population(s) (KQ1 and 2):

- Adults (18 years and over) presenting with symptoms (e.g., an acute episode of joint inflammation) suggestive of gout, including the following subgroups:
  - Male and female patients
  - Older (65 and over) and younger patients
  - Patients with comorbidities including hypertension, type 2 diabetes, kidney disease (renal insufficiency)
  - Patients with osteoarthritis, septic arthritis, or previous joint trauma
  - Individuals with a family history of gout

#### Interventions (KQ1, 2):

- Clinical history and physical exam

- Serum uric acid assessment
- US
- DECT
- Plain x-ray
- Joint aspiration by physicians and synovial fluid analysis using polarizing microscopy (by physicians or laboratory personnel)
- Combinations of these tests as identified in the literature

#### **Comparators:**

- Joint synovial fluid aspiration and microscopic assessment for monosodium urate crystals (KQ1a-c, 2)
- Joint synovial fluid aspiration and microscopic assessment for monosodium urate crystals as performed by a practitioner with a different level of expertise or experience, e.g. rheumatologist, laboratory personnel (KQ1d)

#### **Outcomes:**

- Diagnostic accuracy of clinical signs and symptoms, US, DECT, plain radiographs compared with joint aspiration and synovial fluid analysis (KQ1)
  - sensitivity/specificity, true positives/true negatives, area under the curve
  - positive, negative predictive value, positive/negative likelihood ratios (if prevalence known)
- Clinical decisionmaking
  - Additional testing
  - Pharmacologic/dietary management
- Intermediate outcomes
  - sUA
  - Synovial fluid crystals
  - radiographic or US changes
- Clinical outcomes:
  - pain, joint swelling and tenderness,
  - patient global assessment, and activity limitations (KQ1,2)
- Adverse effects of the tests, including
  - pain, infection, radiation exposure and
  - effects of false positive or false negative (KQ2)

#### **Timing:**

- For clinical outcomes of symptom relief: 1-2 days minimum (KQ1)
- Early in a flare vs. later or post-flare (KQ1c)
- For adverse events: immediate

#### **Settings:**

- Primary care (outpatient) or acute care settings, preferentially;
- Outpatient rheumatology practices/academic medical centers

## **Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions**

The search strategy was designed by the SCEPC reference librarian in collaboration with our local content expert, who has participated in two systematic reviews on gout;<sup>11, 12</sup> the search strategy appears in Appendix A. As recommended by the AHRQ EPC Methods Guide for

Medical Test Reviews,<sup>13</sup> the searches were conducted without filters specific for diagnostic tests; instead we used the terms “gout” combined with the terms for the diagnostic tests. We searched PubMed (1946 to the present), EMBASE (1972 to the present), the Cochrane Collection (1945 to the present), and the Web of Science (from 1949 to the present); these dates were selected to replicate the searches conducted as the basis for the 2006 EULAR Guidelines on Diagnosis and Management of Gout.<sup>14</sup> We also searched Clinicaltrials.gov and the Web of Science for recently completed and other unpublished or non-peer-reviewed studies. Searches were not limited by language of publication; non-English studies that met the inclusion/exclusion criteria based on English abstracts were screened further in full text if translators could be identified with reasonable effort. Manufacturers of diagnostic equipment (polarizing microscopes, sonography equipment, DECT, serum uric acid test kits) were contacted for unpublished data specific to their use for gout diagnosis. Any relevant studies identified for the searches we conducted for a simultaneous review on management of gout were also included if not identified in the searches for this review. Finally, we asked the TEP to assess our included studies and to provide references for any studies they believe should also be included. An update search will be conducted after submission of the draft report for peer review.

The output of the literature searches was transferred to DistillerSR™ for screening. Article titles and abstracts identified by the searches were dually screened by the literature reviewers, and those selected by either reviewer as reporting on diagnosis of gout were accepted without reconciliation for further, full-text review. Full-text review was conducted in duplicate using the predetermined inclusion and exclusion criteria. Disagreements regarding inclusion at the full-text stage were reconciled, with the input of the project lead when necessary. A second round of review was then conducted with full text to exclude articles that provided no usable data, reported duplicate data, or, for studies assessing test validity, did not include or identify control groups (patients who tested negative for gout using the gold standard test). We identified a small number of relatively recent systematic reviews on various aspects of gout diagnosis; in most cases we used these reviews to verify the completeness of our literature searches and as a potential source of additional references, however if the review was of high quality, addressed a subquestion of interest, and included all the literature on the topic, we included it as a data source after assessing its quality. We also searched accepted studies for additional references and screened any articles of apparent interest. For studies of apparent interest reported in meeting abstracts, we searched for peer-reviewed articles before determining whether to accept the studies.

## **Data Abstraction and Data Management**

Articles accepted for inclusion were dually abstracted in DistillerSR, and any disagreements were reconciled with the input of the project leader, SCEPC director, or the local subject matter expert. Included studies went on for dual abstraction of study-level details and outcomes and for assessment of risk of bias. Studies provided by manufacturers or suggested by peer reviewers underwent the same process, as will studies identified in update searches.

## **Assessment of Methodological Quality of Individual Studies**

Risk of bias of individual included studies was assessed in duplicate using QUADAS-2,<sup>15, 16</sup> and assessments were reconciled, with any disagreements mediated by the project lead. We used AMSTAR to assess the quality of existing systematic reviews that we included;<sup>17</sup> AMSTAR assessments were also conducted in duplicate and reconciled.

## **Data Synthesis/Analysis**

For studies that assessed US, DECT, or another radiographic method, we extracted and reported sensitivity, specificity, positive and negative predictive value, and area under the curve (AUC)/receiver-operating characteristics (ROC) if reported, or data provided that allowed us to perform the calculations needed to derive these outcomes.

Studies were considered for meta-analysis if the number of true positives, true negatives, false positives, and false negatives were reported or could be calculated; and if studies were similar enough with respect to outcome measures, participants, and tests, and assessed the validity of an alternative diagnostic method against that of analysis of MSU crystals in synovial fluid. Summary receiver operating characteristic curves and 95% confidence regions were produced, using a bivariate model proposed by Rutter and Gatsonis (2001).<sup>18</sup> Pooled and study level estimates were also plotted. All analyses were conducted in R software V3.1.0.<sup>19</sup>

If data were identified, sensitivity analyses were conducted by age group, sex, particular comorbidities, joint involvement, duration of current symptoms, or type of practitioner. For studies where pooling was not an option, outcomes are described narratively, stratified by test comparisons of interest and study design. All included studies are also described in summary tables.

If any prior SRs were identified that directly addressed a KQ of interest and were deemed of high enough quality to include, we assessed whether any subsequent (or contemporaneous) original studies were sufficiently homogeneous with the review to consider conducting new quantitative synthesis for a particular outcome, based on whether the new study represents a potential pivotal finding in terms of size and effect size and the availability of the needed data. If it was determined that the newer studies could not be combined with the prior SR and could not, themselves, be pooled, we described the newer studies narratively.

## **Grading the Strength of the Body of Evidence for Each Key Question**

We assessed the overall strength of evidence using guidance suggested by AHRQ for its Effective Health Care Program.<sup>20</sup> This method is based loosely on one developed by the GRADE Working Group and classifies the grade of evidence as High, Moderate, Low, or Insufficient. The evidence grade is based on five required domains: study limitations, consistency, directness, precision, and publication bias. Study limitations were assessed using the combined risk-of-bias assessments. Consistency was assessed for the pooled findings based on the 95% confidence intervals; for groups of studies that were not pooled, we assessed the relative sensitivities and specificities. Directness and precision were not assessed. Publication bias was assessed for studies for which data were pooled using the Begg adjusted rank correlation test<sup>21</sup> and Egger regression asymmetry test.<sup>22</sup>

## **Applicability**

Applicability was assessed based on the PICOTS and included the study population age, sex, health profiles (including comorbidities as well as duration of symptoms and number of affected joints, when relevant), tests, gold standards, study settings, and provider types.<sup>23</sup>



# Peer Review and Public Commentary

To Be Added for Final Version

## Results

### Results of Literature Searches

We searched PubMed (1946-April 20, 2014), Web of Science (January 1, 1980-April 21, 2014), Cochrane (earliest through April 21, 2014), Embase (January 1, 1972-present), and SCOPUS (January 1, 2006-April 25, 2014) for peer-reviewed literature. The start dates differ because of the inception dates of the databases being indexed. We also reviewed included studies and excluded but relevant systematic and non-systematic reviews for relevant citations. We also searched Clinicaltrials.gov and Web of Science for unpublished studies and contacted the manufacturers of imaging equipment and test kits used to diagnose gout for results.

Our searches identified 3,473 abstracts, of which 2,989 abstracts were excluded for the following reasons: not human (107), diagnostic methods beyond the scope of the review (129), not gout diagnosis or management (1,702), no original data (125), conference proceedings/presentations/abstracts (11), case reports with less than 10 sample size (408), population under 18 (5), renal transplant/end-stage renal disease (11), no abstracts (247), gout management only (244) (see Figure B). We reviewed 484 full text articles, of which 227 were further excluded for the following reasons: not human (1), diagnostic methods beyond the scope of the review (44), not gout diagnosis or management (69), no original data (29), conference proceedings/presentations/abstracts (37), case reports with less than 10 sample size (17), gout management only (12), no reference standard reported or not all patients received the reference standard (3). We were unable to obtain articles for 15 studies. An additional 235 were potential background articles (non-systematic reviews of potential relevance).

Our search of Clinicaltrials.gov and of the grey literature databases identified no studies of gout diagnosis. None of the manufacturers of imaging equipment or laboratory test kits used in the diagnosis of gout who were contacted for information responded to requests. A notice placed in the Federal Register requesting such information also received no responses.

We include the results of 16 original studies and five systematic reviews in our evidence synthesis. Sixteen studies answer Key Question 1, and two studies answer Key Question 2. Results for these studies can be found below by Key Question. Appendix B contains the list of the studies excluded at full text review, and Appendix D contains our data abstraction tool and QUADAS-2 tools that were used on the 17 included studies and the AMSTAR tool that was used to assess the quality of the included systematic reviews.

**Figure B. Literature flow diagram**

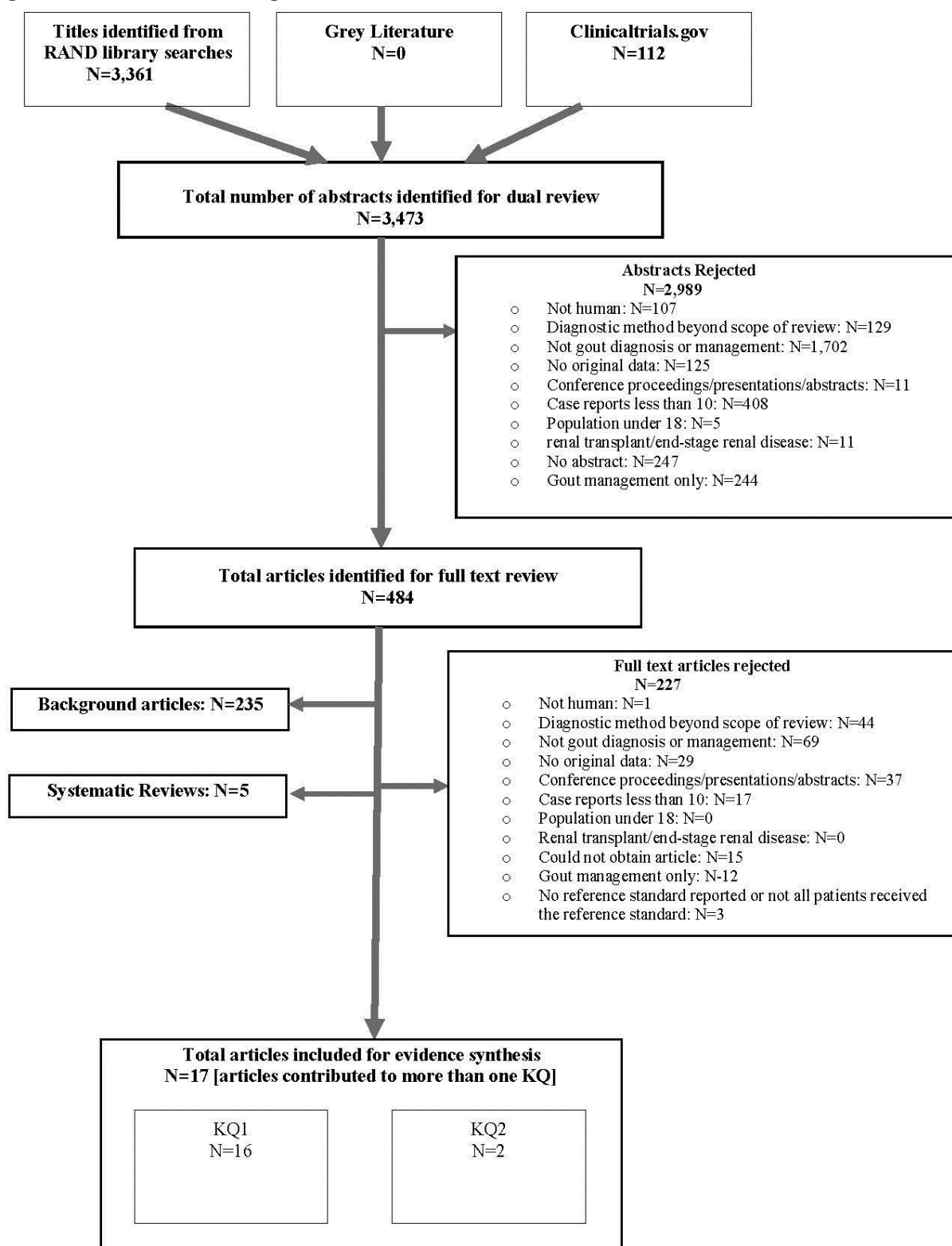


Figure notes: KQ=Key Question

## Findings

The findings of the review are summarized below and in Table B.

### Key Question 1a-c.

- a. What is the accuracy of clinical signs and symptoms and other diagnostic tests (such as serum uric acid, ultrasound, CT scan, DECT, and plain x-ray), alone or in combination, compared to synovial fluid analysis in the diagnosis of acute gouty arthritis, and how does the accuracy affect clinical decision making, clinical outcomes and complications, and patient centered outcomes?
- b. How does the diagnostic accuracy of clinical signs and symptoms and other tests vary by affected joint site and number of joints?
- c. Does the accuracy of diagnostic tests for gout vary by duration of symptoms (i.e., time from the beginning of a flare)

### Description of Included Studies

We identified 15 original studies that met our inclusion criteria for studies on the comparative effectiveness of methods for the diagnosis of gout: 9 studies assessed the sensitivity and specificity of combinations of clinical signs and symptoms (clinical algorithms),<sup>9, 24-31</sup> three assessed the use of DECT,<sup>10, 32, 33</sup> and four assessed the use of US (one study compared US and DECT).<sup>33-36</sup> We also identified four prior systematic reviews: two that assessed the use of imaging for diagnosis of gout,<sup>37, 38</sup> one that examined factors affecting the validity of identification of MSU crystals in synovial fluid,<sup>39</sup> and one on sex differences in gout diagnosis.<sup>40</sup>

The 9 studies that assessed the use of clinical algorithms each compared the predictions based on these tests to assessment of synovial fluid MSU crystals in all or most enrolled patients or at least in those presumed to have gout (in the latter case, patient who were considered not to have gout had to have another condition confirmed by a validated diagnostic criterion). These studies, which dated from 1977, enrolled 82 to 706 adult patients, both male and female. All studies were conducted in academic rheumatology departments, although several of the studies purposefully enrolled patients who were recruited by primary care physicians. Table A compares the components of each of the clinical algorithms.

The three studies that assessed the use of DECT compared the predictions based on these imaging studies to assessment of synovial fluid MSU crystals or to a clinical algorithm or some combination of the two reference standards. These studies dated from 2011 to 2014 and enrolled 31 to 94 patients with suspected gout. All studies were conducted in academic rheumatology departments.

The four studies that assessed the use of US compared the predictions based on US signs to assessment of synovial fluid MSU crystals or to a clinical algorithm or some combination. The studies dated from 2008 to 2014 and enrolled 54 to 105 patients each, with suspected gout.

## Key Points

- Few studies that assessed the accuracy of clinical algorithms consistently applied the same reference standard to all participants with suspected gout and sometimes included participants with chronic gout.
- Studies that assessed the use of clinical algorithms reported widely varying sensitivities and specificities. However, an algorithm developed from clinical signs and symptoms used by primary care physicians reported good positive and negative predictive value. The model with the highest area under the ROC curve (AUC) contained seven predefined variables (male sex, previous patient-reported arthritis attack, onset within 1 day, joint redness, MTP 1 joint involvement, hypertension or one or more cardiovascular diseases, and serum uric acid greater than 5.88mg/dL (AUC 0.85 (95% CI 0.81-0.90); the AUC for the composite without serum uric acid assessment was 0.82). **The strength of evidence for this conclusion is low based on the identification of only two studies that assessed this particular clinical algorithm.**
- In small numbers of studies that enrolled only patients not previously diagnosed with gout, both ultrasound and DECT had good sensitivity and specificity for predicting gout compared with synovial fluid analysis for MSU crystals or a validated clinical algorithm. Three studies revealed sensitivities that ranged from 85% to 100% and specificities that ranged from 83% to 92% for DECT. Four studies of ultrasound (US) showed sensitivities ranging from 37% to 100% and specificities ranging from 68% to 97%, depending on the signs assessed. . **The strength of evidence for this conclusion is low.**
- No studies were identified that assessed the clinical utility of serum uric acid, CT scan, or plain x-ray for diagnosing gout. **The strength of evidence for these tests is insufficient.**
- No studies were identified that directly assessed the effect of joint site or number of affected joints on diagnostic accuracy. **The strength of evidence for this question is insufficient for all diagnostic methods.**
- No studies were identified that directly assessed the effect of duration of symptoms on the accuracy of diagnostic tests. **The strength of evidence for this question is insufficient for all diagnostic methods.**

**Table A. Comparison of Components among Clinical Algorithms for Diagnosis of Gout**

Criteria Components	Wallace 1977 (ARA/ ACR) <sup>9</sup>	Rome <sup>a27</sup>	NY <sup>b27</sup>	EULAR 2006 <sup>14</sup>	Janssens 2010 <sup>28</sup>	CGD 2010 <sup>c41</sup>	3e Initiative 2014 <sup>d42</sup>	Richette Survey <sup>e30</sup>
>1 attack of acute arthritis	✓				✓	✓		✓
Maximum inflammation developed within 1 day	✓	✓ painful joint swelling, abrupt onset, clearing 1-2 wk	✓ 2 attacks	✓	✓	✓		
Monoarthritis/oligoarthritis attack	✓					✓		
Redness observed over joints	✓			✓	✓	✓		
1 <sup>st</sup> MTP joint painful or swollen	✓			✓	✓	✓ podagra	✓ podagra	
Unilateral 1 <sup>st</sup> MTP joint attack	✓		✓					
Unilateral tarsal joint attack	✓					✓		
Tophi (proven or suspected)	✓	✓	✓	✓		✓	✓	✓
Hyperuricemia	✓	✓		✓	✓	✓		✓
Asymmetric swelling within a joint on radiograph	✓			✓				
Subcortical cysts without erosions on radiograph	✓			✓				
Joint fluid culture negative	✓							
MSU crystals in synovial fluid or tissues		✓					✓	
Abrupt onset and remission in 1-2 weeks initially			✓					
Response to colchicine – major reduction in inflammation within 48h			✓				✓	
Male sex					✓			✓
Hypertension or ≥1 CVD					✓			✓

Hypertriglyceridemia								✓
Pain intensity $\geq$ 9/10								✓
Involvement of toes, foot, or ankles								✓
Treatment with corticosteroids								✓
Treatment with NSAIDS								✓
Resolution of pain $<$ 15 days after onset								✓

Table Notes: <sup>a</sup>Meets 2 of the following criteria; <sup>b</sup>MSU crystals in joint fluid or tophus or tissue OR meets 2 of the following criteria; <sup>c</sup> $\geq$ 4/8 of the following criteria; <sup>d</sup>Guideline 1 states that MSU is required for a definitive diagnosis but in its absence, clinical criteria such as the following can be used or characteristic imaging findings may substitute; <sup>e</sup>designed to be administered telephonically by non-physicians to assess prevalence of gout via patient self-report; treatment questions refer to most prominent episode.

## **Key Question 1d. Does the accuracy of synovial fluid aspiration and crystal analysis differ by i) the type of practitioner who is performing the aspiration and ii) the type of practitioner who is performing the crystal analysis?**

### **Description of Included Studies**

We identified one original study that addressed this question directly. We also identified one original study and one medium quality 2013 systematic review that addressed related issues that are also summarized here. A 2014 study was identified that retrospectively audited medical records of two Korean academic medical centers to assess factors associated with false negative synovial fluid MSU results and focused on the personnel performing the analysis and several other factors.<sup>43</sup> A 1987 survey examined the accuracy of MSU and CPPD analysis among hospitals and hospital laboratories in two states.<sup>44</sup> Finally, a 2014 systematic review examined the accuracy of methods for MSU crystal detection in synovial fluid samples and the effects of sample handling.<sup>39</sup>

### **Key Points**

- Detection of MSU crystals in synovial fluid appears to be affected by a number of factors, which include the experience and training of analysts, the storage and handling of samples, and patient factors such as the time lag between the onset of a gout attack and synovial fluid assessment. Because of the relatively small number of studies identified, **the strength of evidence for definitive influential factors is low.**

## **Key Question 2. What are the adverse effects associated with each diagnostic test (including pain, infection at the aspiration site, radiation exposure) or harms (related to false positives, false negatives, indeterminate results) associated with tests used to diagnose gout?**

### **Description of included studies**

One study was identified that assessed adverse effects associated with tests used to diagnose gout.<sup>32</sup> This study reported no adverse events associated with aspiration of synovial fluid for MSU analysis or the use of DECT.

One study examined the outcomes of delayed diagnosis or misdiagnosis of gout in two academic medical centers in South Korea.<sup>43</sup>

### **Key Points**

- No studies documented any adverse events associated with diagnostic tests included in this report. **The strength of evidence for this conclusion is low, based on one study**

**that reported no adverse events associated with joint fluid aspiration for MSU analysis or DECT, and no studies that reported on adverse events associated with US or clinical examination.**

- Misdiagnosis or delayed diagnosis of acute gout can result in unnecessary surgery or hospitalization and delays in adequate treatment, although only one study was identified that examined outpatients with gout. **The strength of evidence for this conclusion is low.** The conclusion is based on one retrospective study in a non-U.S. hospital, but the findings of this study strongly argue for additional studies to be conducted.



**Table B. Summary of Findings and Strength of Evidence**

Key question	Number/type of studies	Strength of evidence	Findings
1a Diagnostic accuracy			
Clinical signs and symptoms (algorithms)	9 observational studies	Low	Tests vary in accuracy compared to synovial fluid aspiration and MSU crystal analysis. One set of criteria based on primary care patients had AUC of 0.85 (95% CI 0.81-0.90)
Dual Emission Computerized Tomography (DECT)	3 observational 1 systematic review	Low	Sensitivity and specificity are good in patients with suspected gout
Ultrasound (US)	4 observational 2 systematic reviews	Low	Sensitivity and specificity are good in patients with suspected gout
Other tests	0 studies	Insufficient	None
1b Influence of number and types of joints involved	0 studies	Insufficient	None
1c Influence of Symptom Duration	0 studies	Insufficient	None
1d. Influence of factors on analysis of monosodium urate crystals (MSU)	2 observational studies 1 systematic review	Low	Results of MSU analysis appear to be affected by a number of factors related to patients (e.g., delayed presentation and joint aspiration), analyst experience, sample handling, and assay.
2. Adverse events and implications of misdiagnosis	2 observational studies: 1 on AEs associated with two diagnostic methods and 1 on implications of misdiagnosis	Low	One study reported DECT and joint aspiration for MSU analysis were associated with no adverse events. One study reported gout misdiagnosis resulted in longer hospital stays, unnecessary surgery, and delayed pharmacological treatment.

## Discussion

### Findings in Relationship to what is Already Known

The development and validation of an algorithm based on clinical signs and symptoms to diagnose gout, as an alternative to the aspiration of synovial fluid and analysis of MSU crystals, has continued since the implementation of the ARA/ACR criteria in the mid-1970s.<sup>9</sup> The present report has attempted to assemble and review all such published algorithms and the attempts at their validation. The challenges to aspiration of synovial fluid and MSU analysis, particularly in the primary care and acute care environment, are well known;<sup>45</sup> thus the need for a valid set of diagnostic tests that can be performed in those settings is clear. The original ARA/ACR criteria and several subsequent modifications, such as the Rome and NY criteria, have demonstrated

limited sensitivity and specificity, and studies aimed at validating them have been criticized for enrolling participants who were hand-selected, often had advanced gout, and other factors that would limit the validity and applicability of the tests in the primary care setting among patients presenting with acute, often initial attacks of gout.<sup>24 3</sup> In addition, attempts to simplify diagnostic criteria to a few elements such as serum uric acid and response to colchicine have been shown not to be helpful.<sup>46</sup>

As a result of the lack of strong specificity of the various algorithms, the 2006 EULAR diagnostic criteria for gout,<sup>14</sup> based on a literature review and expert panel consensus, asserted that clinical diagnosis was not definitive without identification of MSU in synovial fluid or tophus. Synovial fluid MSU analysis was recommended in patients with no prior diagnosis of gout and in asymptomatic patients during intercritical periods.<sup>47</sup> Because septic arthritis must be ruled out, the guidelines also advocate culturing synovial fluid if septic arthritis is suspected (and both conditions can coexist). However, implementation of these guidelines has clearly remained challenging in the primary care setting, as evidenced by further research and recent efforts to set new guidelines for diagnosis in the primary care setting.<sup>28</sup> This Diagnostic Rule has been shown to perform better than the ACR criteria in comparable populations.<sup>29</sup>

The 2011 Postgraduate Medicine guidelines for diagnosis of gout (which aimed to update the EULAR 2006 guidelines) also emphasize reliance on clinical signs and symptoms, while acknowledging that MSU constitutes the definitive diagnosis.<sup>48</sup> Signs and symptoms include acute monoarticular attacks of the lower extremities; rapid development of severe pain, swelling, and tenderness that peaks within 6 to 12 hours, especially with overlying erythema is thought to be highly indicative of crystal formation, but not necessarily MSU crystals. The Postgraduate Medicine guidelines also note that diagnosis based on clinical signs and symptoms alone has reasonable accuracy when patients have typical presentation of gout, and that sUA levels cannot be used to diagnose or rule out gout. They also confirm the need for culture if sepsis is suspected. No studies were identified that assessed the accuracy of this algorithm.

Current effort appears to be focused on refining the Diagnostic Rule described above and on new classification criteria for research on gout.<sup>49</sup> In addition, the 3e (Evidence, Expertise, Exchange) initiative (3ei), a multinational effort to promote evidence-based practice, issued a set of recommendations in 2014 on the diagnosis and treatment of gout, combining a literature review and expert opinion.<sup>42</sup> The 3ei recommendation on gout diagnosis recognizes the use of MSU as the gold standard but also the difficulty in performing this test under some circumstances, asserting that if MSU cannot be performed, the diagnosis “can be supported by classical clinical features (such as podagra, tophi, rapid response to colchicine), and/or characteristic imaging findings.” They note the poor diagnostic utility of most of the clinical criteria, by themselves; however, they make an exception for response to colchicine and the presence of tophi, although the former has been demonstrated to be unable to definitely distinguish gout from other crystal arthritis forms (and the latter is often seen only in more advanced gout).

## **Accuracy of DECT for the Diagnosis of Gout**

DECT is a non-invasive study method that can detect urate deposits in joints, tendons, bursa, and soft tissues. The radiographic signature of urate can be distinguished from that of calcium. DECT requires special machines and software to process the images and currently is not widely available. Radiation exposure is not greater than standard CT scanning and is limited to extremities, which are not radio-sensitive organs.

Studies assessing the diagnostic utility of DECT are promising, demonstrating high sensitivity and specificity for gout. However, we identified only a small number of studies on patients without previous diagnoses of gout.

Bongartz' recent (2014) publication<sup>32</sup> sought to determine the additive value of DECT to a clinically unclear presentation among 30 patients. Of these 30, 14 had a positive DECT, and 11 of 12 (2 patients refused aspiration) had crystal confirmation of gout using ultrasound guided aspiration. All 4 patients with false negative DECT from another group of 40 patients with confirmed gout, had new onset gout (first attack and symptom duration < 6 months). Glazebrook had also prospectively studied inflammatory mono-arthritis patients, demonstrating high sensitivity and specificity for crystal-confirmed gout cases.<sup>10</sup>

The summary of the literature demonstrates that DECT is both specific and sensitive for gout. Utility of DECT may be best for evaluating urate burden in established gout patients. Limited data suggest that for patients with recurrent attacks of inflammatory mono- or oligo-arthritis where the question of gout is unresolved (for example, no fluid available for aspiration or negative study), DECT should demonstrate good diagnostic value. However, for patients with a first inflammatory mono-articular attack (due to gout), DECT may not be sensitive. The availability of DECT machines in most regions also may limit application of this technology.

## **Accuracy of Ultrasound for the Diagnosis of Gout**

Although we identified only a small number of studies assessing the accuracy of US for diagnosis of gout in patients without a previous diagnosis, its use as a diagnostic test appears promising. Sensitivity and specificity for specific findings or amalgamation of findings were typically high. In addition, it is relatively inexpensive, non-invasive, and well accepted by patients.

However, several challenges must be overcome prior to this being accepted as a standard. There are several ultrasound characteristics of gout including the “double contour sign”, characteristic intra-articular findings (bright spots or “snow”), or tophaceous findings. As follows the nature of the disease, these findings can present in many different joints and the analyses we reviewed each used different methodology for identifying which joints were studied. The number of joints studied ranged from a single target (inflamed) joint up to 26 joints. Additionally, up to 20 tendon areas and 6 bursae were also examined. Exhaustive scanning is not practical. Some authors (notably Lamers-Karnebeck)<sup>34</sup> described limited systematic evaluation of inflammatory mono-arthritis patients with sensitivities and likelihood ratios for specific findings. Nevertheless, this focused methodology (4 to 6 joints) is beyond what would be available from most radiology centers, which typically focus on more comprehensive examinations of single joints. With most of these publications in rheumatology journals or by rheumatology authors, diagnostic enthusiasm for gout and such studies appears to be greatest in the rheumatology community. Furthermore, we did not find studies that evaluated the marginal utility of ultrasound data beyond clinical criteria or in lieu of joint aspiration.

Thus, the present review also confirms the results of several relatively recent systematic reviews on the validity and potential superiority of DECT and ultrasound for the diagnosis of gout. However, as the 3e Recommendations note, the “availability, cost, and the need for trained personnel and specific equipment...” might limit their use in routine clinical practice. Thus, these guidelines seem to suggest that in primary care settings, diagnosis can be based on a set of clinical criteria.<sup>49</sup>

## **Applicability**

Two factors may reduce the applicability of this review:

1. In many of the studies that assessed diagnostic algorithms, participants had already had a definitive diagnosis of gout, and sometimes had chronic gout. Relatively few of these studies enrolled only participants with suspected gout or monoarthritis. Three of the five studies included in the pooled analysis of the accuracy of DECT and three of the four studies included in the pooled analysis of US enrolled only gout suspects.
2. All studies were conducted in a rheumatology setting, usually an academic rheumatology department. Patients being seen in this setting may have more advanced disease than those seen in a primary care setting. A small number of studies purposefully recruited patients with PCP-diagnosed or suspected gout, expressly to assess the validity of PCP clinical decision making in gout diagnosis.

## **Implications for Clinical and Policy Decisionmaking**

The findings of this review provide some evidence to support the further development and validation of diagnostic algorithms based on a combination of clinical signs and symptoms for the diagnosis of gout in the primary care setting, with the use of imaging modalities (US and DECT) in cases where a definitive diagnosis cannot be made from signs and symptoms alone.

## **Limitations of the Comparative Effectiveness Review Process**

Assessing the comparative validity of diagnostic tests in systematic reviews presents a number of challenges that are not faced with comparative effectiveness reviews of treatment strategies. These limitations are magnified by several issues surrounding tests for gout and the natural history of the disease itself. To increase applicability, we limited included studies to those that enrolled previously undiagnosed patients; in doing so, we excluded a number of studies on the use of ultrasound and DECT for monitoring gout. Previous systematic reviews on the use of US and DECT included studies of patients with asymptomatic hyperuricemia, studies of patients with suspected gout but without definitive diagnoses, and studies of patients with definitive gout diagnoses in various stages of the disease.

In addition, our consideration of unpublished literature was limited. Although the Scientific Resource Center requested information from manufacturers of microscopes and imaging equipment used to diagnose gout (directly and through a notice in the Federal Register), no information was provided; we did not search FDA databases for such information ourselves. In addition, we included only conference proceedings cited in other included studies or suggested by TEP members or the subject matter expert on our project team.

## **Limitations of the Evidence Base**

The literature that addresses the diagnosis of gout has numerous limitations that make it difficult to draw firm conclusions. These limitations can be divided into three categories: study volume, design, and reporting quality.

## **Volume**

In short, few studies have attempted to address the diagnosis of gout. Almost no studies have examined the impact of diagnostic test accuracy on decision-making (decisions to order further testing or to initiate particular treatments) or any clinical or patient centered outcomes, and almost no studies addressed adverse events potentially associated with diagnostic testing. Most studies of gout address management issues or monitoring of patients with chronic gout.

## **Design**

Of the diagnostic studies we did identify, almost no studies of diagnostic algorithms limited enrollment to gout suspects or patients with a monoarthritis or some other clinical signs or symptoms that might suggest gout. Many studies enrolled patients with known gout, and many included no control group.

Even studies that enrolled patients who were gout suspects or included a control group systematically failed to limit enrollment to patients in their first attack or with recent onset or did not stratify findings by duration of the condition (as would be ascertained by asking, “How long have you been having these attacks?”). The lack of stratification by duration of condition would likely affect the positive and negative predictive value of imaging techniques more than it would affect diagnostic tests based on clinical signs and symptoms, but not necessarily, as one criterion in the latter is almost always the presence of tophus.

Most studies also fail to stratify by other relevant factors, such as time since the onset of the current or most recent flare, sex, and comorbidities. The time since onset of the current flare definitely affects the presence of crystals as well as clinical signs and symptoms.

No studies tested the validity of a diagnostic algorithm comprising clinical signs and symptoms and an imaging test, compared with a clinical algorithm or imaging alone.

Finally, issues concerning the use of synovial fluid MSU crystal identification as the reference standard abound. Taking the validity of the reference standard at face value, some studies assessed MSU in a fraction of participants only (e.g., those for whom synovial fluid could be aspirated, those most suspected of having gout, or those willing to undergo the test), using the ACR criteria or individual clinical judgment as the reference standard for the remaining participants. The technical problems with aspiration and analysis have been assessed and described extensively and include inconsistencies introduced by patient factors (e.g., the time lapse from the start of the flare to aspiration), sample handling factors (storage duration and temperature), and practitioner skills in aspiration and analysis.

## **Reporting Quality**

Failure to report important study design details in publications is a further limitation. Studies tended to be vague regarding blinding of assessors and the time lapse between implementation of the index test and reference standard (and the sequence of tests), a critical detail considering the short duration of gout attacks.

## Research Gaps

As described in the section on the limitations of the research base, above, promising algorithms have been validated in an insufficient number of studies, particularly studies in primary care settings.

Patient-level factors that influence test behavior have also been understudied: These include the influence of duration of a flare; number and identity of joints involved; and patient age, sex, and comorbidities.

Finally, studies may be needed to emphasize the impact of misdiagnosis of gout, either failure to diagnose gout and misdiagnosis as septic or osteoarthritis or failure to diagnose conditions such as septic arthritis.

## Conclusions

This review highlights the need to validate promising diagnostic algorithms in primary care settings. An algorithm with high diagnostic accuracy can ideally form part of a decision tree that combines clinical signs and symptoms with more invasive tests or imaging for clinically ambiguous cases.

## References

1. Li C, Martin BC, Cummins DF, et al. Ambulatory resource utilization and cost for gout in United States. *American Journal of Pharmacy Benefits*. 2013 March/April;5(2):e46-e54. PMID: 2013264278.
2. Doghramji PP, Wortmann RL. Hyperuricemia and gout: new concepts in diagnosis and management. *Postgrad Med*. 2012 Nov;124(6):98-109. PMID: 23322143.
3. Wallace SL. Hyperuricemia in the diagnosis of gout. *J Gen Intern Med*. 1989 Mar-Apr;4(2):178-9. PMID: 2651605.
4. Singh JA, Reddy SG, Kundukulam J. Risk factors for gout and prevention: a systematic review of the literature. *Curr Opin Rheumatol*. 2011 Mar;23(2):192-202. PMID: 21285714.
5. Hamburger M, Baraf HS, Adamson TC, 3rd, et al. 2011 Recommendations for the diagnosis and management of gout and hyperuricemia. *Postgrad Med*. 2011 Nov;123(6 Suppl 1):3-36. PMID: 22156509.
6. Harrold LR, Mazor KM, Negron A, et al. Primary care providers' knowledge, beliefs and treatment practices for gout: results of a physician questionnaire. *Rheumatology (Oxford)*. 2013 Sep;52(9):1623-9. PMID: 23620554.
7. Owens D, Whelan B, McCarthy G. A survey of the management of gout in primary care. *Irish Medical Journal*. 2008;101(5).
8. Perez-Ruiz F, Carmona L, Yébenes MJG, et al. An Audit of the Variability of Diagnosis and Management of Gout in the Rheumatology Setting The Gout Evaluation and Management Study. *Jcr-Journal of Clinical Rheumatology*. 2011 Oct;17(7):349-55. PMID: WOS:000295320000001.
9. Wallace SL, Robinson H, Masi AT, et al. Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum*. 1977 Apr;20(3):895-900. PMID: 856219.
10. Glazebrook KN, Guimaraes LS, Murthy NS, et al. Identification of intraarticular and periarticular uric acid crystals with dual-energy CT: initial evaluation. *Radiology*. 2011 Nov;261(2):516-24. PMID: 21926378.
11. Khanna D, Fitzgerald JD, Khanna PP, et al. 2012 American college of rheumatology guidelines for management of gout. part 1: Systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. *Arthritis Care and Research*. 2012;64(10):1431-46.
12. Khanna D, Khanna PP, Fitzgerald JD, et al. 2012 American college of rheumatology guidelines for management of gout. part 2: Therapy and antiinflammatory prophylaxis of acute gouty arthritis. *Arthritis Care and Research*. 2012;64(10):1447-61.
13. Whiting P, Westwood M, Beynon R, et al. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol*. 2011 Jun;64(6):602-7. PMID: 21075596.
14. Zhang W, Doherty M, Pascual E, et al. EULAR evidence based recommendations for gout. Part I: Diagnosis. Report of a task force of the Standing Committee for

International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis*. 2006 Oct;65(10):1301-11. PMID: 16707533.

15. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011 Oct 18;155(8):529-36. PMID: 22007046.

16. Santaguida PL, Riley CM, Matchar DB. Assessing Risk of Bias as a Domain of Quality in Medical Test Studies. In: Chang SM, Matchar DB, Smetana GW, Umscheid CA, eds. *Methods Guide for Medical Test Reviews*. Rockville (MD); 2012.

17. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10. PMID: 17302989.

18. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001 Oct 15;20(19):2865-84. PMID: 11568945.

19. Team RDC. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>. 2014.

20. Chang SM, Matchar DB, Smetana GW, et al. *Methods Guide for Medical Test Reviews*. AHRQ Publication No. 12-EC017 Agency for Healthcare Research and Quality. Rockville (MD): 2012. <http://www.ncbi.nlm.nih.gov/pubmed/22834019>.

21. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994 Dec;50(4):1088-101. PMID: 7786990.

22. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997 Sep 13;315(7109):629-34. PMID: 9310563.

23. Atkins D, Chang S, Gartlehner G, et al. *Assessing the Applicability of Studies When Comparing Medical Interventions*. Agency for Healthcare Research and Quality; December 2010. *Methods Guide for Comparative Effectiveness Reviews*. AHRQ Publication No. 11-EHC019-EF.

24. Kienhorst L, Janssens H, Fransen J, et al. The performance of a diagnostic rule for acute gouty arthritis without joint fluid analysis: A validation study. *Arthritis and Rheumatism*. 2013 October;65 SUPPL. 10:S500.

25. Lenski M, Scherer MA. Analysis of synovial inflammatory markers to differ infectious from gouty arthritis. *Clin Biochem*. 2014 Jan;47(1-2):49-55. PMID: 24177196.

26. Kienhorst LB, Janssens HJ, Fransen J, et al. Arthritis of the first metatarsophalangeal joint is not always gout: A prospective cohort study in primary care patients. *Joint Bone Spine*. 2014 Jan 24; PMID: 24468668.

27. Malik A, Schumacher HR, Dinnella JE, et al. Clinical diagnostic criteria for gout: comparison with the gold standard of synovial fluid crystal analysis. *J Clin Rheumatol*. 2009 Feb;15(1):22-4. PMID: 19125136.



28. Janssens HJEM, Fransen J, Van De Lisdonk EH, et al. A diagnostic rule for acute gouty arthritis in primary care without joint fluid analysis. *Archives of Internal Medicine*. 2010;170(13):1120-6.
29. Janssens HJEM, Janssen M, Van De Lisdonk EH, et al. Limited validity of the American college of rheumatology criteria for classifying patients with gout in primary care. *Annals of the Rheumatic Diseases*. 2010;69(6):1255-6.
30. Richette P, Clerson P, Bouee S, et al. Identification of patients with gout: Elaboration of a questionnaire for epidemiological studies. *Annals of the Rheumatic Diseases*. 2014.
31. Vazquez-Mellado J, Hernandez-Cuevas CB, Alvarez-Hernandez E, et al. The diagnostic value of the proposal for clinical gout diagnosis (CGD). *Clin Rheumatol*. 2012 Mar;31(3):429-34. PMID: 21979446.
32. Bongartz T, Glazebrook KN, Kavros SJ, et al. Dual-energy CT for the diagnosis of gout: an accuracy and diagnostic yield study. *Ann Rheum Dis*. 2014 Mar 25 PMID: 24671771.
33. Huppertz A, Hermann KG, Diekhoff T, et al. Systemic staging for urate crystal deposits with dual-energy CT and ultrasound in patients with suspected gout. *Rheumatol Int*. 2014 Mar 12 PMID: 24619560.
34. Lamers-Karnebeek FB, Van Riel PL, Jansen TL. Additive value for ultrasonographic signal in a screening algorithm for patients presenting with acute mono-/oligoarthritis in whom gout is suspected. *Clin Rheumatol*. 2014 Apr;33(4):555-9. PMID: 24510062.
35. Rettenbacher T, Ennemoser S, Weirich H, et al. Diagnostic imaging of gout: comparison of high-resolution US versus conventional X-ray. *Eur Radiol*. 2008 Mar;18(3):621-30. PMID: 17994238.
36. Lai KL, Chiu YM. Role of Ultrasonography in Diagnosing Gouty Arthritis. *Journal of Medical Ultrasound*. 2011 March;19(1):7-13. PMID: 2011227256  
FULL TEXT LINK  
<http://dx.doi.org/10.1016/j.jmu.2011.01.003>.
37. Ogdie A, Taylor W, Weatherall M, et al. Imaging modalities for the classification of gout: Systematic literature review and meta-analysis. *Annals of the Rheumatic Diseases*. 2014;73.
38. Chowalloor PV, Keen HI. A systematic review of ultrasonography in gout and asymptomatic hyperuricaemia. *Ann Rheum Dis*. 2013 May;72(5):638-45. PMID: 23291387.
39. Graf SW, Buchbinder R, Zochling J, et al. The accuracy of methods for urate crystal detection in synovial fluid and the effect of sample handling: a systematic review. *Clin Rheumatol*. 2013 Feb;32(2):225-32. PMID: 23138881.
40. Dirken-Heukensfeldt K, Teunissen TAM, van de Lisdonk EH, et al. "Clinical features of women with gout arthritis." A systematic review. *Clinical Rheumatology*. 2010 Jun;29(6):575-82. PMID: WOS:000277096500001.
41. Pelaez-Ballestas I, Cuevas CH, Burgos-Vargas R, et al. Diagnosis of Chronic Gout: Evaluating the American College of Rheumatology Proposal, European League Against Rheumatism Recommendations, and Clinical Judgment. *Journal of*

Rheumatology. 2010 Aug;37(8):1743-8. PMID: WOS:000280860600028.

42. Sivera F, Andres M, Carmona L, et al. Multinational evidence-based recommendations for the diagnosis and management of gout: integrating systematic literature review and expert opinion of a broad panel of rheumatologists in the 3e initiative. *Annals of the Rheumatic Diseases*. 2014 Feb;73(2):328-35. PMID: WOS:000329488000001.

43. Park JW, Ko DJ, Yoo JJ, et al. Clinical factors and treatment outcomes associated with failure in the detection of urate crystal in patients with acute gouty arthritis. *Korean Journal of Internal Medicine*. 2014;29(3):361-9.

44. Hasselbacher P. Variation in synovial fluid analysis by hospital laboratories. *Arthritis and Rheumatism*. 1987;30(6):637-42.

45. Segal JB, Albert D. Diagnosis of crystal-induced arthritis by synovial fluid examination for crystals: lessons from an imperfect test. *Arthritis Care Res*. 1999 Dec;12(6):376-80. PMID: 11081008.

46. Rigby AS, Wood PH. Serum uric acid levels and gout: what does this herald for the population? *Clin Exp Rheumatol*. 1994 Jul-Aug;12(4):395-400. PMID: 7955603.

47. Chen LX, Schumacher HR. Gout: can we create an evidence-based systematic approach to diagnosis and management? *Best Practice and Research: Clinical Rheumatology*. 2006 August;20(4):673-84. PMID: 2006438744 MEDLINE PMID 16979531  
(<http://www.ncbi.nlm.nih.gov/pubmed/16979531>) FULL TEXT LINK  
<http://dx.doi.org/10.1016/j.berh.2006.05.006>

48. Hamburger MI, Tesser JRP, Skosey JL, et al. Patterns of gout treatment and related outcomes in us community rheumatology practices: The relation between gout flares, time in treatment, serum uric acid level and urate lowering therapy. *Arthritis and Rheumatism*. 2012 October;64 SUPPL. 10:S808-S9.

49. Dalbeth N. Management of gout in primary care: challenges and potential solutions. *Rheumatology (Oxford)*. 2013 Sep;52(9):1549-50. PMID: 23771950.



# Introduction

## Background

### Condition

Gout is a form of inflammatory arthritis characterized by acute intermittent episodes of synovitis presenting with joint swelling and pain (referred to as acute gouty arthritis) that may progress to a chronic intermittent condition, which may progress further to development of tophi (solid deposits of monosodium urate [MSU] crystals in joints, cartilage, and bones), a condition called chronic tophaceous gout.

Based on data from the 2007-2008 National Health and Nutrition Examination Survey (NHANES), the prevalence of gout among adults in the United States has been estimated to be 3.9 percent (8.3 million individuals), ranging from 2.0 percent in women to 5.9 percent in men.<sup>1</sup> Comparing the most recent figures for the prevalence of gout to those of previous cycles of NHANES shows that the prevalence of gout appears to be increasing. The rise in the prevalence of gout has paralleled the increase in prevalence of conditions associated with hyperuricemia, including obesity, hypertension, hypertriglyceridemia, hypercholesterolemia, type 2 diabetes and metabolic syndrome, chronic kidney disease, and renal insufficiency. Certain medications also may increase the risk for developing gout (e.g., thiazide diuretics).

A 2013 study estimated the annual costs of gout to be \$933 million (in 2008 figures), with the annual ambulatory care costs associated with gout potentially reaching \$1 billion. Some 32 percent of the costs were attributed to gouty arthritis attacks, and drug expenditures accounted for 61 percent of the total costs.<sup>1</sup>

**Etiology of Gout.** The driving force behind acute episodes of gout is hyperuricemia (defined as a serum uric acid (strictly, urate, sUA) concentration greater than 6.8 mg per deciliter [dl] in men and greater than 6.0 in women). Hyperuricemia can be the result of either inadequate renal excretion of UA or, less commonly, UA overproduction (UA is a breakdown product of dietary or endogenous purines, which are among the building blocks of nucleic acids); and is associated with the formation and deposition of the UA crystals, which preferentially dissolve, in joints, tendons, and bursa spaces. Despite the prevalence of hyperuricemia, for reasons that remain unclear, only a small proportion of individuals with hyperuricemia go on to develop gout; in the rest, hyperuricemia remains asymptomatic.<sup>2</sup> The prevalence of hyperuricemia ranges from 21.2 percent in men to 21.6 percent in women, four- to ten-fold higher than the prevalence of gout.<sup>3</sup>

The causes of gout are multifactorial, including a combination of genetic, hormonal, metabolic, pharmacologic, renal disease, and dietary factors. Family history, advancing age, male sex, or, in women, early menopause have been associated with a higher risk of gout and/or gout flares.<sup>4</sup> Dietary risk factors for gout appear to include alcohol consumption, as well as consumption of meat, seafood, sugar sweetened soft drinks, and foods high in fructose, whereas dairy foods and coffee have been associated with a lower risk of incident gout and in some cases a lower rate of gout flares. However, the role of diet in the etiology and treatment of gout is a topic of considerable research and will be reviewed in a separate systematic review.

## Diagnosis Strategies

Some research has supported the need for laboratory assessment of joint/synovial fluid MSU in the setting of an acute inflammatory arthritis for a definitive diagnosis of gout, a test generally conducted by a rheumatologist to confirm a conditional/presumptive diagnosis.<sup>5</sup> However, the majority of individuals with gout are initially seen, diagnosed, and treated in primary and urgent care settings. Thus primary care physicians (PCPs) and emergency medicine physicians are the most likely practitioners to see patients with early-stage gout (with more advanced or difficult-to-diagnose patients likely being referred to rheumatologists). However, use of the gold standard synovial fluid analysis for diagnosis of gout is difficult and seldom performed in the primary or emergent care setting.<sup>6, 7</sup> In fact, evidence from a 2011 survey of rheumatologists suggests that SF analysis is underused in the rheumatology setting as well.<sup>8</sup> Instead, PCPs and emergency medicine physicians may tend to rely on an algorithm comprising a combination of clinical signs and symptoms to diagnose an acute episode of gout. Attempts to standardize and validate such diagnostic algorithms date back to the 1970s.<sup>9</sup> A question of interest is whether any combination of clinical signs and symptoms and laboratory tests accessible in the primary or acute care setting will have good predictive value compared with tests such as joint aspiration and synovial fluid analysis. Plain radiographs have also been used for gout diagnosis, and newer techniques, including ultrasound (US) and dual-energy computed tomography (DECT) are just beginning to be used to diagnose gout in some settings.<sup>10</sup> Another question of interest is how these newer methods compare with joint aspiration and synovial fluid analysis in their predictive value for the diagnosis of gout.

Therefore, we have undertaken a systematic review of studies examining the accuracy of tests used to diagnose gout, including combinations of physical findings, serum UA, US, plain radiography, and DECT, compared with synovial fluid UA. The results of this review should help inform clinical decision-making for patients and providers and improve the quality of care for patients with gout in the primary and acute care setting.

## Scope and Key Questions

### Scope of the Review

The purpose of this review is to assess the evidence on the comparative validity and safety of tests used for the diagnosis of gout, including clinical signs and symptoms (individually and in combination as an algorithm), DECT, US, and other visualization methods, compared with the gold standard of aspiration of synovial fluid from involved joints and analysis of monosodium urate (MSU) crystals using polarized light microscopy. The review also assesses the evidence that practitioner type and other factors may affect the outcomes of MSU analysis. AHRQ assigned this report to the Southern CA Evidence-based Practice Center (HHS A290201200006I). A protocol for the review was posted on the AHRQ website on July 17, 2014 at: <http://effectivehealthcare.ahrq.gov/ehc/products/564/1937/gout-protocol-140716.pdf>. The protocol was approved by the AHRQ Center for Evidence and Practice Improvement.

## Key Questions

Figure 1 shows an analytic framework to illustrate the population, interventions, outcomes, and adverse effects that will guide the literature search and synthesis for this project.

**Figure 1. Analytic Framework**

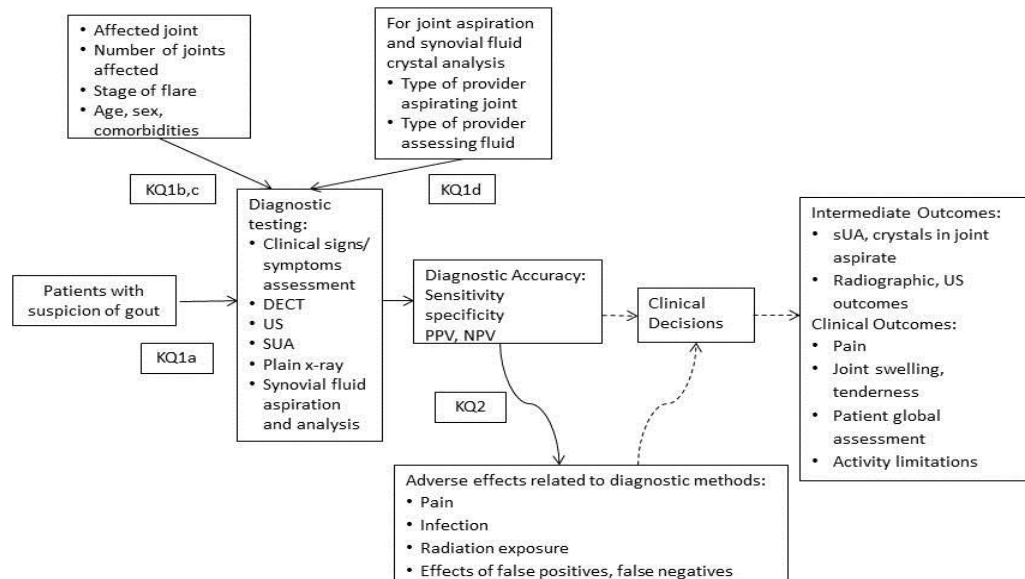


Figure notes: KQ=Key Question; NPV=Negative predictive value; PPV=Positive predictive value; sUA=Serum uric acid; US=Ultrasound

## The Key Questions

The key questions that guided this review are based on questions posed by the ACP. These questions underwent revision based on input from a group of key informants, public comments, and input from a Technical Expert Panel (TEP).

### Key Question 1.

- What is the accuracy of clinical signs and symptoms and other diagnostic tests (such as serum uric acid, ultrasound, CT scan, DECT, and plain x-ray), alone or in combination, compared to synovial fluid analysis in the diagnosis of acute gouty arthritis, and how does the accuracy affect clinical decision making, clinical outcomes and complications, and patient centered outcomes?
- How does the diagnostic accuracy of clinical signs and symptoms and other tests vary by affected joint site and number of joints?
- Does the accuracy of diagnostic tests for gout vary by duration of symptoms (i.e., time from the beginning of a flare)
- Does the accuracy of synovial fluid aspiration and crystal analysis differ by i) the type of practitioner who is performing the aspiration and ii) the type of practitioner who is performing the crystal analysis?

## Key Question 2.

What are the adverse effects associated with each diagnostic test (including pain, infection at the aspiration site, radiation exposure) or harms (related to false positives, false negatives, indeterminate results) associated with tests used to diagnose gout?

## Organization of this report

The remainder of this report presents the methods used to conduct the literature searches, data abstraction, and analysis for this review; the results of the literature searches; the conclusions; and a discussion of the limitations as well as suggestions for future research.

# Methods

## Criteria for Inclusion/Exclusion of Studies in the Review

This report is based on a systematic search for prospective or cross-sectional studies that compared the sensitivity/specificity of tests used to diagnose gout against the gold standard test of joint aspiration and synovial fluid assessment for MSU crystals in populations of adults 18 years of age or older, suspected of having gout (see Table 1, below). Tests of interest included clinical examination for physical signs, symptoms, and history (in the form of algorithms); serum uric acid, ultrasonography; DECT; and plain radiography. The comparator of primary interest was synovial fluid analysis of MSU crystals using polarized light microscopy. Studies were also included if some or all of the index test comparisons were to the American College of Rheumatology criteria for gout diagnosis as a reference standard (comparator). Studies were excluded if enrolled participants had already been definitively diagnosed with gout or if the comparator was individual physician opinion or was not identified.

Outcomes of interest were the accuracy of the test results (the sensitivity and specificity or the positive and negative predictive value of the test in question), intermediate outcomes such as lab and radiographic test results, clinical decision making, short term clinical (patient-centered) outcomes such as pain and joint swelling, and any adverse events (including adverse patient experiences such as pain or infection at the aspiration site, effects of radiation exposure, and the results of a false positive or negative) associated with the test. Studies were also accepted if they examined factors that potentially affected the validity of tests (including but not limited to joints involved, duration of symptoms, sex, or types of practitioners performing or evaluating the tests).

Prospective cohort, cross-sectional, and case-control (if needed) studies that compared the sensitivity/specificity or area under the curve between a proposed diagnostic test and the gold standard were included to address KQ1a. Studies of similar design that compared outcomes based on the joint(s) involved (KQ1b), the lag between the onset of symptoms and diagnosis (commencement of treatment) (KQ1c) or the type of practitioner who conducted or analyzed the test (or other similar factors affecting test results) were also included. Prospective cohort, case control, and case series of any size, as well as case reports of rare adverse events were included if they addressed adverse events or other negative outcomes in individuals undergoing testing (KQ2).

## PICOTS

### Population(s) (KQ1 and 2):

- Adults (18 years and over) presenting with symptoms (e.g., an acute episode of joint inflammation) suggestive of gout, including the following subgroups:
  - Male and female patients
  - Older (65 and over) and younger patients
  - Patients with comorbidities including hypertension, type 2 diabetes, kidney disease (renal insufficiency)
  - Patients with osteoarthritis, septic arthritis, or previous joint trauma
  - Individuals with a family history of gout

### Interventions (KQ1, 2):

- Clinical history and physical exam



- Serum uric acid assessment
- US
- DECT
- Plain x-ray
- Joint aspiration by physicians and synovial fluid analysis using polarizing microscopy (by physicians or laboratory personnel)
- Combinations of these tests as identified in the literature

#### **Comparators:**

- Joint synovial fluid aspiration and microscopic assessment for monosodium urate crystals (KQ1a-c, 2)
- Joint synovial fluid aspiration and microscopic assessment for monosodium urate crystals as performed by a practitioner with a different level of expertise or experience, e.g. rheumatologist, laboratory personnel (KQ1d)

#### **Outcomes:**

- Diagnostic accuracy of clinical signs and symptoms, US, DECT, plain radiographs compared with joint aspiration and synovial fluid analysis (KQ1)
  - sensitivity/specificity, true positives/true negatives, area under the curve
  - positive, negative predictive value, positive/negative likelihood ratios (if prevalence known)
- Clinical decisionmaking
  - Additional testing
  - Pharmacologic/dietary management
- Intermediate outcomes
  - sUA
  - Synovial fluid crystals
  - radiographic or US changes
- Clinical outcomes:
  - pain, joint swelling and tenderness,
  - patient global assessment, and activity limitations (KQ1,2)
- Adverse effects of the tests, including
  - pain, infection, radiation exposure and
  - effects of false positive or false negative (KQ2)

#### **Timing:**

- For clinical outcomes of symptom relief: 1-2 days minimum (KQ1)
- Early in a flare vs. later or post-flare (KQ1c)
- For adverse events: immediate

#### **Settings:**

- Primary care (outpatient) or acute care settings, preferentially;
- Outpatient rheumatology practices/academic medical centers

## **Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions**

The search strategy was designed by the SCEPC reference librarian in collaboration with our local content expert, who has participated in two systematic reviews on gout;<sup>11, 12</sup> the search strategy appears in Appendix A. As recommended by the AHRQ EPC Methods Guide for

Medical Test Reviews,<sup>13</sup> the searches were conducted without filters specific for diagnostic tests; instead we used the terms “gout” combined with the terms for the diagnostic tests. We searched PubMed (1946 to the present), EMBASE (1972 to the present), the Cochrane Collection (1945 to the present), and the Web of Science (from 1949 to the present); these dates were selected to replicate the searches conducted as the basis for the 2006 EULAR Guidelines on Diagnosis and Management of Gout.<sup>14</sup> We also searched Clinicaltrials.gov and the Web of Science for recently completed and other unpublished or non-peer-reviewed studies. Searches were not limited by language of publication; non-English studies that met the inclusion/exclusion criteria based on English abstracts were screened further in full text if translators could be identified with reasonable effort. Manufacturers of diagnostic equipment (polarizing microscopes, sonography equipment, DECT, serum uric acid test kits) were contacted for unpublished data specific to their use for gout diagnosis. Any relevant studies identified for the searches we conducted for a simultaneous review on management of gout were also included if not identified in the searches for this review. Finally, we asked the TEP to assess our included studies and to provide references for any studies they believe should also be included. An update search will be conducted after submission of the draft report for peer review.

The output of the literature searches was transferred to DistillerSR™ for screening. Article titles and abstracts identified by the searches were dually screened by the literature reviewers, and those selected by either reviewer as reporting on diagnosis of gout were accepted without reconciliation for further, full-text review. Full-text review was conducted in duplicate using the predetermined inclusion and exclusion criteria. Disagreements regarding inclusion at the full-text stage were reconciled, with the input of the project lead when necessary. A second round of review was then conducted with full text to exclude articles that provided no usable data, reported duplicate data, or, for studies assessing test validity, did not include or identify control groups (patients who tested negative for gout using the gold standard test). We identified a small number of relatively recent systematic reviews on various aspects of gout diagnosis; in most cases we used these reviews to verify the completeness of our literature searches and as a potential source of additional references, however if the review was of high quality, addressed a subquestion of interest, and included all the literature on the topic, we included it as a data source after assessing its quality and how it assessed risk of bias and strength of evidence. We also searched accepted studies for additional references and screened any articles of apparent interest. For studies of apparent interest reported in meeting abstracts, we searched for peer-reviewed articles before determining whether to accept the studies.

## **Data Abstraction and Data Management**

Articles accepted for inclusion were dually abstracted in DistillerSR, and any disagreements were reconciled with the input of the project leader, SCEPC director, or the local subject matter expert. Included studies went on for dual abstraction of study-level details and outcomes and for assessment of risk of bias. Studies provided by manufacturers or suggested by peer reviewers underwent the same process, as will studies identified in update searches.

## **Assessment of Methodological Quality of Individual Studies**

Risk of bias of individual included studies was assessed in duplicate using QUADAS-2,<sup>15, 16</sup> and assessments were reconciled, with any disagreements mediated by the project lead. We used

AMSTAR to assess the quality of existing systematic reviews that we included;<sup>17</sup> AMSTAR assessments were also conducted in duplicate and reconciled.

## **Data Synthesis/Analysis**

For studies that assessed US, DECT, or another radiographic method, we extracted and reported sensitivity, specificity, positive and negative predictive value, and area under the curve (AUC)/receiver-operating characteristics (ROC) if reported, or data provided that allowed us to perform the calculations needed to derive these outcomes.

Studies were considered for meta-analysis if the number of true positives, true negatives, false positives, and false negatives were reported or could be calculated; and if studies were similar enough with respect to outcome measures, participants, and tests, and assessed the validity of an alternative diagnostic method against that of analysis of MSU crystals in synovial fluid. Summary receiver operating characteristic curves and 95% confidence regions were produced, using a bivariate model proposed by Rutter and Gatsonis (2001).<sup>18</sup> Pooled and study level estimates were also plotted. All analyses were conducted in R software V3.1.0.<sup>19</sup>

If data were identified, sensitivity analyses were conducted by age group, sex, particular comorbidities, joint involvement, duration of current symptoms, or type of practitioner. For studies where pooling was not an option, outcomes are described narratively, stratified by test comparisons of interest and study design. All included studies are also described in summary tables.

If any prior SRs were identified that directly addressed a KQ of interest and were deemed of high enough quality to include, we assessed whether any subsequent (or contemporaneous) original studies were sufficiently homogeneous with the review to consider conducting new quantitative synthesis for a particular outcome, based on whether the new study represents a potential pivotal finding in terms of size and effect size and the availability of the needed data. If it was determined that the newer studies could not be combined with the prior SR and could not, themselves, be pooled, we described the newer studies narratively.

## **Grading the Strength of the Body of Evidence for Each Key Question**

We assessed the overall strength of evidence using guidance suggested by AHRQ for its Effective Health Care Program.<sup>20</sup> This method is based loosely on one developed by the GRADE Working Group and classifies the grade of evidence as High, Moderate, Low, or Insufficient. The evidence grade is based on five required domains: study limitations, consistency, directness, precision, and publication bias. Study limitations were assessed using the combined risk-of-bias assessments. Consistency was assessed for the pooled findings based on the 95% confidence intervals; for groups of studies that were not pooled, we assessed the relative sensitivities and specificities. Directness and precision were not assessed. Publication bias was assessed for studies for which data were pooled using the Begg adjusted rank correlation test<sup>21</sup> and Egger regression asymmetry test.<sup>22</sup>

## **Applicability**

Applicability was assessed based on the PICOTS and included the study population age, sex, health profiles (including comorbidities as well as duration of symptoms and number of affected joints, when relevant), tests, gold standards, study settings, and provider types.<sup>23</sup>

# **Peer Review and Public Commentary**

To Be Added for Final Version

# Results

## Introduction

This chapter first describes the results of the literature searches and then provides the results for each key question, including key points, an overview of the studies identified for that question, and a detailed synthesis of the studies. The results of Key Question 1, parts a through c, are presented together for each diagnostic method. The results for Key Question 1d are described following that section.

## Results of Literature Searches

We searched PubMed (1946-April 20, 2014), Web of Science (January 1, 1980-April 21, 2014), Cochrane (earliest through April 21, 2014), Embase (January 1, 1972-present), and SCOPUS (January 1, 2006-April 25, 2014) for peer-reviewed literature. The start dates differ because of the inception dates of the databases being indexed. We also reviewed included studies and excluded but relevant systematic and non-systematic reviews for relevant citations. We also searched Clinicaltrials.gov and Web of Science for unpublished studies and contacted the manufacturers of imaging equipment and test kits used to diagnose gout for results.

Our searches identified 3,473 abstracts, of which 2,989 abstracts were excluded for the following reasons: not human (107), diagnostic methods beyond the scope of the review (129), not gout diagnosis or management (1,702), no original data (125), conference proceedings/presentations/abstracts (11), case reports with less than 10 sample size (408), population under 18 (5), renal transplant/end-stage renal disease (11), no abstracts (247), gout management only (244) (see Figure 2). We reviewed 484 full text articles, of which 227 were further excluded for the following reasons: not human (1), diagnostic methods beyond the scope of the review (44), not gout diagnosis or management (69), no original data (29), conference proceedings/presentations/abstracts (37), case reports with less than 10 sample size (17), gout management only (12), no reference standard reported or not all patients received the reference standard (3). We were unable to obtain articles for 15 studies. An additional 235 were potential background articles (non-systematic reviews of potential relevance).

Our search of Clinicaltrials.gov and of the grey literature databases identified no studies of gout diagnosis. None of the manufacturers of imaging equipment or laboratory test kits used in the diagnosis of gout who were contacted for information responded to requests. A notice placed in the Federal Register requesting such information also received no responses.

We include the results of 17 original studies and five systematic reviews in our evidence synthesis. Sixteen of the original studies answer Key Question 1, and 2 studies answer Key Question 2. Results for these studies can be found below by Key Question. Appendix B contains the list of the studies excluded at full text review, and Appendix D contains our data abstraction tool and QUADAS-2 tools that were used on the included studies and the AMSTAR tool that was used to assess the quality of the included systematic reviews.

**Figure 2. Literature flow diagram**

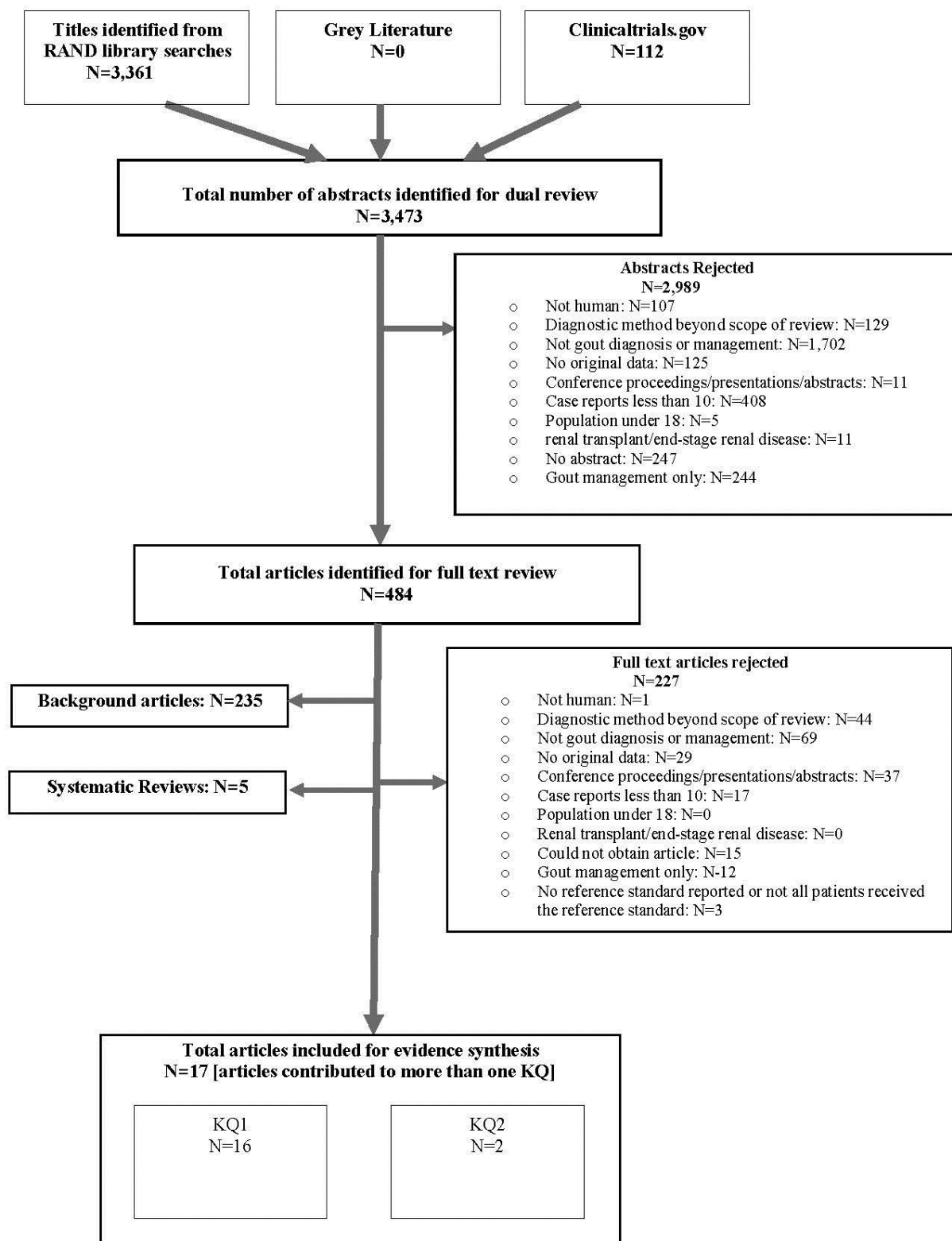


Figure notes: KQ=Key Question

### Key Question 1a-c:

- a. What is the accuracy of clinical signs and symptoms and other diagnostic tests (such as serum uric acid, ultrasound, CT scan, DECT, and plain x-ray), alone or in combination, compared to synovial fluid analysis in the diagnosis of acute gouty arthritis, and how does the accuracy affect clinical decision making, clinical outcomes and complications, and patient centered outcomes?
- b. How does the diagnostic accuracy of clinical signs and symptoms and other tests vary by affected joint site and number of joints?
- c. Does the accuracy of diagnostic tests for gout vary by duration of symptoms (i.e., time from the beginning of a flare)

### Key Points

- Few studies that assessed the accuracy of algorithms comprising clinical signs and symptoms consistently applied the same reference standard to all participants with suspected gout and sometimes included participants with chronic gout.
- Studies that assessed the use of clinical algorithms reported widely varying sensitivities and specificities; however, an algorithm developed from clinical signs and symptoms used by primary care physicians reported good positive and negative predictive value and was validated in a small secondary care population but needs further validation. **The strength of evidence for this conclusion is low.**
- In small numbers of studies that enrolled only patients not previously diagnosed with gout, both ultrasound and DECT had good sensitivity and specificity for predicting gout compared with synovial fluid analysis for MSU crystals or a validated clinical algorithm. Three studies revealed sensitivities that ranged from 85% to 100% and specificities that ranged from 83% to 92% for DECT. Four studies of US showed sensitivities that ranged from 37% to 100% and specificities ranging from 68% to 97%, depending on the signs assessed. **The strength of evidence for this conclusion is low.**
- No studies were identified that assessed the clinical utility of serum uric acid, CT scan, or plain x-ray for diagnosing gout. **The strength of evidence for these tests is insufficient.**
- No studies were identified that directly assessed the effect of joint site or number of affected joints on diagnostic accuracy. **The strength of evidence for this question is insufficient for all diagnostic methods.**
- No studies were identified that directly assessed the effect of duration of symptoms on the accuracy of diagnostic tests. **The strength of evidence for this question is insufficient for all diagnostic methods.**

### Description of included studies

We identified 15 original studies that met our inclusion criteria for studies on the comparative effectiveness of methods for the diagnosis of gout: 9 studies assessed the sensitivity and specificity of algorithms comprising combinations of clinical signs and symptoms,<sup>9, 24-31</sup> three assessed the use of DECT,<sup>10, 32, 33</sup> and four assessed the use of US (one study compared US and DECT).<sup>33-36</sup> We also identified four prior systematic reviews: two that assessed the use of

imaging for diagnosis of gout,<sup>37,38</sup> one that examined factors affecting the validity of identification of MSU crystals in synovial fluid,<sup>39</sup> and one on sex differences in gout diagnosis.<sup>40</sup>

The 9 studies that assessed the use of clinical algorithms each compared the predictions based on these tests to assessment of synovial fluid MSU crystals in all or most enrolled patients or at least in those presumed to have gout (in the latter case, patient who were considered not to have gout had to have another condition confirmed by a validated diagnostic criterion). These studies, which dated from 1977, enrolled 82 to 706 adult patients, both male and female. All studies were conducted in academic rheumatology departments, although several of the studies purposefully enrolled patients who were recruited by primary care physicians. Table 1 describes the studies and Table 2 compares the components of each algorithm.

The three studies that assessed the use of DECT compared the predictions based on these imaging studies to assessment of synovial fluid MSU crystals or to a clinical algorithm or some combination of the two reference standards. These studies dated from 2011 to 2014 and enrolled 31 to 94 patients with suspected gout. All studies were conducted in academic rheumatology departments.

The four studies that assessed the use of US compared the predictions based on these imaging studies to assessment of synovial fluid MSU crystals or to a clinical algorithm or some combination of the two reference standards. These studies dated from 2008 to 2014 and enrolled 54 to 105 patients with monoarthritis or suspected gout.

## **Detailed Synthesis**

This section describes the included studies and the findings for each diagnostic method. In cases where we were able to pool studies for a meta-analysis, we describe the overall study characteristics and results.

## **Studies Assessing Algorithms Comprising Clinical Signs and Symptoms to Diagnose Gout**

### **Comparative Accuracy**

The nine studies that comparatively assessed the use of algorithms comprising clinical signs and symptoms of gout in diagnosis are described in Table 1. Table 2 compares the criteria assessed in the studies as well as additional guideline-based criteria. The QUADAS scores for risk of bias for each of the included studies are shown in Table 5 at the end of this chapter.

In 1977, the American Rheumatism Association (ARA, forerunner of the American College of Rheumatology) Subcommittee on Classification Criteria for Gout, compiled a survey questionnaire comprising 53 proposed criteria for classifying acute gout, both for diagnosis and clinical decision making, for epidemiology, and for research.<sup>9</sup> Thirty eight rheumatologists from 38 academic and private rheumatology clinics throughout the United States tested the criteria on 706 patients (average age: 61): 178 patients provisionally diagnosed with gout, 299 with rheumatoid arthritis (RA), and the remainder with acute septic arthritis or pseudogout. The proportion of females was 13.6% for gout patients, and ranged from 51 to 64% for the other disorders. Among the 178 patients presumed to have gout, 90 underwent joint aspiration; of those, 85% had positive findings. The 53 criteria included in the survey were narrowed to the 12 that best discriminated the study groups and showed the greatest balance of sensitivity (the proportion of MSU-positive patients who fulfilled the proposed criterion) and specificity (the



proportion of control patients who did not fulfill the proposed criterion) (Table 1). The sensitivity of three different cutpoints—5 or more, 6 or more, and 7 or more positive criteria—were 95.5%, 87.6%, and 74.1%, respectively. The specificities for these cutpoints were 72.7%-93.3%, 89.1%-98.7%, and 97.3%-100%, respectively; in each case, the criteria showed the poorest discrimination between gout and pseudogout.

In 2009, researchers enrolled 82 patients (mean age 64.5, 6% female) at a Veterans Administration rheumatology clinic to compare the ARA criteria<sup>9</sup> with two subsequent, briefer variations, designated the New York Criteria and the Rome Criteria (Table 2).<sup>27</sup> The criterion for enrollment was having had synovial fluid aspirated and analyzed for MSU crystals (by two highly experienced rheumatologists) at some prior time; participants were surveyed but did not receive new physical exams for the study. Of the participants, 75.6% had undergone aspiration of a knee joint; the remainder had had the MTP, wrist, elbow, ankle, or proximal interphalangeal joint aspirated. Thirty patients were MSU+. Considering only the clinical signs and symptoms in the three sets of criteria (i.e., excluding MSU crystal findings), none of the clinical criteria sets had a sensitivity of more than 70.0% or a specificity of more than 88.5%. Having two of the three (non-MSU) Rome criteria had the highest specificity (88.55) and PPV (76.9%). Assessing the LR+ of individual criteria, they found that proven or suspected tophus and response to colchicine within 48 hours had the highest LR+ (15.56 [95%CI 2.11, 114.71] and 4.33[95%CI 1.16, 16.16], respectively) and PPV (91% and 86%, respectively) of any of the criteria in the three criteria sets.

In 2010, Janssens published the findings of several studies specifically focused on patients identified in primary care settings.<sup>28, 29</sup> From 2004 to 2007, family physicians in the Netherlands consecutively recruited 381 patients with monoarthritis; the patients were sent, along with their blinded PCP diagnosis (the index test), to an academic rheumatology clinic for MSU crystal analysis, clinical exam, and followup of at least 1 year. Of the 381, 328 patients (mean age 58.0; 20.4% female) had a PCP diagnosis of gout. The researchers analyzed the patients' MSU to validate the PCP diagnoses and used the patients' clinical signs and symptoms, along with lab values (serum UA), to develop a set of multivariate predictive models (six clinical signs and symptoms, with and without the addition of sUA testing). The PCPs' diagnosis had a PPV of 0.64 and a NPV of 0.87. The model with the highest AUC (0.85, 95% CI 0.81, 0.90) contains the elements shown in Figure 2. Avoidance of sUA testing decreased the AUC only slightly (to 0.82).

Janssen' group also utilized the population of 381 PCP-referred patients with monoarthritis and a positive or negative PCP diagnosis of gout to validate the ARA/ACR criteria<sup>9</sup> against the presence of MSU crystals in synovial fluid.<sup>29</sup> Increasing the cutoff points from 4 or more positive criteria (out of 11) to a maximum of 9 or more decreased the sensitivity from 1.00 to 0.07 and increased the specificity from 0.24 to 0.99. The cutoff point of 6 or more criteria showed a sensitivity of 0.80, specificity of 0.64, PPV of 0.80 and NPV of 0.65. In comparison to this study, which enrolled a primary care population, Wallace's preliminary assessment of the criteria, which enrolled a cohort of rheumatology patients, showed a sensitivity of 87.6% and a range of specificities of 89.1%-98.7%.<sup>9</sup>

In 2010, Vazquez-Mellado and colleagues developed a set of criteria that could be used to diagnose chronic gout in the primary care setting without reliance on joint fluid analysis of MSU, the Clinical Gout Diagnosis (CGD) criteria, and then sought to validate the criteria on a population of patients suspected of having acute gout.<sup>31</sup> They recruited 192 consecutive patients (mean age 45; 24% of gout patients were female, compared with 90% of OA patients) seen in a

Mexico City academic rheumatology clinic; 75 had MSU+ gout, and the remainder had RA, osteoarthritis (OA), or spondyloarthritis (SpA). The index test was the 8-item CGD (see Table 2); all patients were assessed for synovial fluid MSU. At a cutoff point of 4 or more positive items out of 8, the sensitivity was 97.3% and the specificity was 95.6%; the LRP was 22.1 and the LRN was 0.03. Compared with the overall patient cohort, the mean number of positive criteria fulfilled by female gout patients was 5.22(SD 1.35) and the mean number for patients with younger onset of gout was 7.0(1.22). All patients with early-onset gout and 94.4% of female gout patients fulfilled 4 or more of the CGD criteria.

In 2013, Janssens' group conducted additional analysis of their 2010 dataset to assess the sensitivity and specificity of arthritis of the MTP1 joint as a diagnostic criterion for gout, as this clinical sign was often the sole criterion used by Dutch PCPs to diagnose and initiate treatment of gout.<sup>26</sup> The population comprised 159 of the initially recruited 381 patients, who had monoarthritis of MTP1. All patients underwent MSU-crystal assessment: At baseline, 74.2% of the patients had crystals; the remainder were followed for at least 1 year; during this followup period, 5 additional patients tested positive for MSU crystals (19 died during the followup period). The PCP diagnosis based on MTP1 monoarthritis had a sensitivity of 99%, a specificity of 8%, a PPV of 0.79, and a NPV of 0.75. The authors concluded that this clinical sign was a promising predictor of gout but not entirely reliable (as 25% of patients were incorrectly diagnosed with gout).

Janssens' group also validated their criteria set on a group of 390 patients seen in a rheumatology clinic for the first time between 2011 and 2013 with signs and symptoms suggestive of gout.<sup>24</sup> As reported in an abstract presented at the 2013 American College of Rheumatology/Association of Rheumatology Health Professionals Annual Meeting, of the 390 patients, 56% had a positive finding of MSU crystals. At cutoff scores of 4 or less positive criteria, 4 to 7 positive criteria, and 8 or more positive criteria, the prevalence of a positive MSU finding was 6%, 46% and 88%, respectively. The AUC was 0.85 and the Hosmer-Lemeshow goodness of fit was 0.64, suggesting that the criteria set developed for use in a primary care population works equally well in the rheumatology setting.

In 2014, a group of French rheumatologists published the results of their attempt to develop and validate a telephone survey/questionnaire to be used by non-medical personnel to elicit self-reports of gout for epidemiological studies (most immediately, to establish the prevalence of gout in France).<sup>30</sup> They recruited a cohort of 246 patients with arthritis who had undergone synovial fluid aspiration and analysis (102 had MSU crystal-confirmed gout and 142 had no MSU crystals) from 14 rheumatology departments throughout France (patients were age 18 or over, mean age 60, 15.7% females among gout patients and 58.5% without gout). The questionnaire contained 62 items; based on the odds ratios of individual items for predicting gout, the researchers developed two logistic regression models, one with 9 of the most predictive items and one with 5 items, and a classification and regression tree (CART) model that contained the 11 items included in the two regression models (Table 2). As shown in Table 1, the three models had similar performance, correctly classifying 90.0%, 88.8%, and 88.5% of patients.

A 2014 study conducted in Germany took a different approach to distinguishing infectious arthritis from gout.<sup>25</sup> The study, conducted in the Department of Trauma Surgery and Orthopedics of an academic teaching hospital in Dachau, analyzed multiple inflammatory markers in serum and synovial fluid drawn from 82 patients (mean age 72.4; 30% females among gout patients; 53 with septic arthritis (47% female) and 29 with gout(30% females)) in the emergency department from 2008 to 2012. Gout and septic arthritis were ascertained by

synovial fluid aspiration with MSU crystal identification and culture, respectively. Among the markers assayed (e.g., serum UA, synovial fluid white blood cells, synovial fluid total protein), synovial fluid lactate had the greatest diagnostic potential to differentiate septic arthritis from gout, followed by glucose and serum uric acid concentrations (Table 1).

### **The Effect of Test Accuracy on Clinical Decision Making, Clinical Outcomes and Complications, and Patient Centered Outcomes**

No studies were identified that tested the effects of the accuracy of gout diagnoses based on clinical signs and symptoms on clinical decision making (treatment decisions or decisions to conduct further testing), clinical outcomes, or patient centered outcomes.

### **The Effect of Affected Joint Site(s) and Number of Joints on the Diagnostic Accuracy of Clinical Signs and Symptoms**

No studies were identified that directly tested the effects of the joint site affected or the number of joints affected on the accuracy of diagnostic tests based on clinical signs and symptoms. One study described above assessed the accuracy of a PCP diagnosis based on MTP1 arthritis compared with synovial fluid analysis of MSU.<sup>26</sup> The PCP diagnosis based on MTP1 monoarthritis had a sensitivity of 99%, a specificity of 8%, a PPV of 0.79, and a NPV of 0.75. The authors concluded that this clinical sign was a promising predictor of gout but not entirely reliable (as 25% of patients were incorrectly diagnosed with gout).

### **The Effect of Duration of Symptoms on the Accuracy of Gout Diagnoses Based on Clinical Signs and Symptoms**

No studies tested the effect of duration of symptoms (time since beginning of flare) on the accuracy of gout diagnoses based on clinical signs and symptoms.

**Table 1. Summary of Included Studies of Algorithms Comprising Clinical Signs and Symptoms Used to Diagnose Gout**

<b>Author, Year Country Practice setting</b>	<b>Population n, putative diagnosis, mean age, % female</b>	<b>Test Components/Reference Standard</b>	<b>Outcome Measures and Findings</b>	<b>Conclusions and Other Comments</b>
Wallace 1977 <sup>9</sup> U.S. 38 academic or private rheumatology practices	706 patients (178 gout; remainder pseudogout, RA, acute septic arthritis) Mean age: 61 % female: 13.6 (gout) (51-64% remainder )	ARA criteria: Maximum inflammation 1 day >1 attack Monoarticular arthritis Redness 1 <sup>st</sup> MTP pain or swelling Unilateral first MTP Unilateral tarsal Suspected tophus Hyperuricemia Asymmetric swelling Subcortical cysts, no erosions Culture negative Reference: positive join MSU, proven tophi, or survey findings	Sensitivity, specificity of increasing numbers of criteria  50.6% of gout patients had joint aspiration (compared with 25% of RA patients, 82.7% of pseudogout patients, 23.7% RA patients, and 70.6% of SA patients); Of those gout patients who underwent MSU, 85% were positive. Presence of tophi had 99% specificity but not absolute. ≥5 +Clinical criteria: Sensitivity: 95.5% Specificity: 72.7-93.3% ≥6+criteria: Sensitivity: 87.6% Specificity: 89.1%-98.7% (pseudogout and SA, respectively) ≥7+criteria: Sensitivity: 74.1% Specificity: 97.3%-100%	A score of 6 or more among the 12 ARA criteria showed high sensitivity and specificity for predicting gout and discriminating from pseudogout, RA, and SA
Malik 2009 <sup>27</sup> US VA rheumatology clinic [Include]	82 patients (37% MSU+) Mean age 64.5 6% female	ARA Criteria: See Wallace '77 NY Criteria: MSU crystals in joint fluid or tissue or tophus OR meets 2 of the following criteria: 2 attacks of painful limb joint swelling Abrupt onset and remission in 1–2 wk initially First MTP attack Presence of a tophus Response to colchicine- major reduction in	Sensitivity and specificity, PPV 6/12 ARA: Sensitivity: 70.0% Specificity: 78.8% PPV:65.6%  2 of 4 NY Criteria: Sensitivity: 70.0% Specificity: 82.7% PPV:70.0%  2 of 3 Rome: Sensitivity: 66.7% Specificity: 88.5% PPV:76.9%	Considering only the clinical elements in the 3 sets of criteria (i.e., not MSU crystal presence), no sets of clinical criteria were more than 70.0% sensitive or 88.5% specific. Having 2 of the 3 Rome criteria had the highest PPV of 76.9%. Authors conclude MSU analysis should remain gold standard.

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Test Components/Reference Standard	Outcome Measures and Findings	Conclusions and Other Comments
		inflammation within 48 h Rome Criteria: Meets 2 of the following criteria: Painful joint swelling, abrupt onset, Clearing in 1–2 wk initially Serum uric acid: 7 in males, 6 in Females Presence of tophi Presence of urate crystals (MSU) in synovial fluid or tissues Reference: SF MSU crystal analysis	(Table 3 of Malik reports the sensitivity, specificity, PPV and likelihood ratio for all ARA criteria and others)	
Janssens 2010a <sup>28</sup> Netherlands Primary care clinics	328 patients with monoarthritis (but otherwise blinded PCP diagnosis) Mean age: 58.0 (13.) years, % female: 20.4% Recruited consecutively from primary care clinics with	Family physician diagnosis, patient-level factors (e.g., age, sex, CVD, onset within one day or hour, family history, medication use; Table 1 of Janssens includes all variables considered) Reference: SF MSU crystal analysis	PPV, NPV, AUC Family physician diagnosis: PPV: 0.64 NPV: 0.87 3 multivariate models were constructed. The most appropriate model contained the following predefined variables: Male sex Previous patient-reported arthritis attack Onset within 1 day Joint redness MTP 1 joint involvement Hypertension or 1 or more CVD, and sUA > 5.88mg/dL* AUC 0.85 (95% CI 0.81-0.90).	Performance did not change after transforming the regression coefficients to easy-to-use scores and was almost equal to that of the statistically optimal model (AUC 0.87; 95% CI, 0.83-0.91)
Janssens 2010b <sup>29</sup> Netherlands	381 monoarthritis patients recruited from primary care clinics (328 with PCP diagnosis of gout) (same population as Janssens 2010a)	ARA/ACR criteria Reference Standard: SF MSU crystal analysis	<u>ARA/ACR criteria (≥4 +)</u> Sensitivity: 1.00, Specificity: 0.24 PPV: 0.70 NPV: 0.97 Overall fraction correct: 72% <u>ARA/ACR criteria (≥6 +)</u> Sensitivity: 0.80 Specificity: 0.64 PPV: 0.80	Increasing stringency of ARA criteria beyond the cutoff point of 6 or more criteria did not improve usability of the criteria greatly: Patients referred from primary care with suspected gout (monoarthritis) assessed with MSU showed that the ARA/ACR criteria not particularly useful in diagnosis.

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Test Components/Reference Standard	Outcome Measures and Findings	Conclusions and Other Comments
			NPV: 0.65 Overall fraction correct: 74% <u>ARA/ACR criteria (<math>\geq 9</math> +)</u> Sensitivity: 0.07 (15/209) Specificity: 0.99 (118/119) PPV: 0.94 (15/16) NPV: 0.38 (118/312) Overall fraction correct: 41% (133/328)	
Vazquez-Mellado 2012 <sup>31</sup> Mexico Academic rheumatology departments	192 consecutive patients with diagnoses of RA, OA, SpA, or gout (75) Mean age: 45 % female: 24% (gout); 90% (OA)	Clinical gout diagnostic [CGD] criteria= $\geq 4/8$ of the following: Current/past history of $>1$ attack of acute arthritis Mono- or oligoarthritis Rapid progression of pain/swelling ( $<24$ hours) Podagra Erythema Unilateral tarsitis Tophi Hyperuricemia Reference standard: MSU crystal analysis performed in all gout patients	Sensitivity, specificity, AUC, PPV, positive likelihood ratio of CGD criteria and effects of patient characteristics (obesity, HTN, CVD, dyslipidemia, hyperglycemia, metabolic syndrome, chronic renal failure) <u><math>\geq 3</math> CGD items:</u> Sensitivity: 99.1 Specificity: 79.8 +LR: 4.90 -LR: 0.01 PPV: 85.6 NPV: 48.7 AUC: 0.90 <u><math>\geq 4</math> CGD items:</u> Sensitivity: 97.3 (95% CI 90.8, 99.3) Specificity: 95.6 (89.2, 98.3) +LR: 22.1 -LR: 0.03 PPV: 94.8 NPV: 97.8 AUC: 0.96	Authors state that study is limited by setting (rheumatology clinic vs. primary care setting) and the patient population likely to be seen (chronic vs. acute) but high concordance with other criteria, e.g., ACR suggests these criteria may be useful in primary care settings. Radiographic data not included because not specific.  The mean CGD criteria in women: 5.22 (SD1.35) The mean CGD criteria in early-onset gout: 7.0(1.22) The kappa value comparing the CGD and the Janssens criteria was 0.85.
Kienhorst 2013 <sup>24</sup> Netherlands Academic rheumatology department	390 patients with suspected gout referred to rheumatology clinic by PCPs	Janssens diagnostic criteria <sup>28</sup> : Male sex Previous patient-reported arthritis attack, Onset within one day Joint redness, MTP 1 involvement, HTN or $\geq 1$ CVD	Prevalence of gout by cut-off scores: $\leq 4$ : 6% (4 of 69) $>4$ to $<8$ : 46% (76 of 163) $\geq 8$ : 88% (138 of 157) AUC: 0.85	Diagnostic criteria developed for primary care settings validated in secondary care population

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Test Components/Reference Standard	Outcome Measures and Findings	Conclusions and Other Comments
		sUA >0.35 mmol/L (5.88 mg/dL) Reference standard: MSU crystal analysis performed in all patients		
Kienhorst 2014 <sup>26</sup> Netherlands Academic rheumatology department	159 primary care patients recruited by PCPs and referred to academic rheumatology department for suspected gout based on MTP1 monoarthritis (of the initially recruited 381PCP patients with suspected gout) Mean age 58.3 % female: 15% confirmed gout, 47% non-gout	PCP diagnosis, primarily based on arthritis of MTP 1 joint, at baseline and, in MSU-, at followup of at least 1 year Reference standard: MSU crystal analysis performed in all patients	Sensitivity, specificity, PPV, NPV of PCP diagnosis at baseline and after followup of >1 year Additional differences in clinical signs and symptoms between gout and non-gout patients  At baseline, 74.2% of patients were MSU+; During followup, 5 additional patients tested MSU+. 19 (11.9%) died during follow-up. PCP diagnosis: Sensitivity: 0.99 Specificity: 0.08 PPV: 0.79 NPV: 0.75 Positive likelihood ratio: 1.1 Negative likelihood ratio: 0.1	Clinical signs and symptoms significantly increased in MSU+ patients by univariate analysis: male sex, previous arthritis, patient-reported gout, use of diuretics, use of CVD or antiHTN drugs, hypertension and/or CVD, beer consumption, BMI> 25, sUA>349umol/L, serum creatinine>105.2umol/L, GFR<60ml/min, CRP>1mg/dL (all p<0.05)  Gout is a prominent cause but not the only cause of MTP-1 monoarthritis, which was incorrectly diagnosed as gout in 25% of patients. These additional signs and symptoms increase precision of diagnosis.
Richette 2014 <sup>30</sup> France Academic rheumatology departments	246 patients (102 with MSU-confirmed gout and 142 MSU-minus controls) from 14 rheumatology departments who agreed to complete a questionnaire Mean age 60 % female: 15.7% confirmed gout, 58.5% without gout	62-item phone questionnaire designed to be administered by non-medical personnel to elicit self-reported clinical signs and symptoms of gout (sociodemographic variables, comorbidities, and characteristics of the most prominent episode of gout) Reference standard: MSU crystal analysis	OR for item responses predicting cases of gout, as determined using multi-adjusted models. Sensitivity, specificity, Positive LR, negative LR, AUC for the logistic regression models and for self-reported gout.  Model 1 (8 items): Sensitivity: 88.0% (79.6, 93.4) Specificity: 93% (87, 96.4) PLR: 12.5 (6.8, 22.8) NLR: 0.13 (0.08, 0.22) Correctly classified 90% (86.4, 94.1) AUC: 0.97 (0.95, 0.99)  Model 2 (5 items):	The 11-item questionnaire correctly classified gout patients more accurately than self-report. This questionnaire will be used to estimate the prevalence of gout in France.

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Test Components/Reference Standard	Outcome Measures and Findings	Conclusions and Other Comments
			Sensitivity: 87.5% (79.6, 93.4) Specificity: 89.8% (82.9, 94.3) PLR: 8.6 (5.1, 14.5) NLR: 0.14 (0.08, 0.24) Correctly classified 88.8% (83.8, 94.192.5) AUC: 0.95 (0.92, 0.98)  Classification and regression tree (CART): Sensitivity: 81.3% (72.2, 88.1) Specificity: 93.7% (88.0, 96.9) PLR: 12.8(6.8, 24.3) NLR: 0.20 (0.13, 0.30) Correctly classified 88.5% (83.7, 92.1) AUC:0.87(0.83, 0.91)	
Lenski 2014 <sup>25</sup> Germany Academic trauma surgery and orthopedics department	82 patients, 29 with gout and 53 with culture verified septic arthritis Mean age: 72.4 % female: 30% for gout patients, 47% for septic arthritis	Retrospective assessment of serum markers (lactate, glucose, uric acid, LDH, SFWBC, total protein Reference standard: MSU crystal analysis for gout patients	Sensitivity, specificity, PPV, NPV, AUC, LR+, LR- Serum uric acid: Sensitivity: 70(52.1, 83.3) Specificity: 85.2(67.5, 94.1) PPV: 71.7 NPV: 83.4 LR+: 4.73 LR-: 0.35 AUC 0.822(0.706, 0.937)  Serum lactate: Sensitivity: 89.5(75.9, 95.8) Specificity: 77.3(56.6, 89.9) LR+: 3.94 LR-:0.14 AUC 0.901(0.823, 0.979)	Synovial lactate had the greatest diagnostic potential to discriminate septic arthritis from gout, followed by glucose and uric acid.

Table Notes: ACR American College of Rheumatology; ARA American Rheumatism Association; AUC area under the curve; BMI body mass index; CGD Clinical Gout Diagnosis; CI confidence interval; CRP c-reactive protein; CVD cardiovascular disease; GFR glomerular filtration rate; HTN hypertension; LR likelihood ratio; MSU monosodium urate; MTP metatarsophalangeal; OA osteoarthritis; NPV negative predictive value; NY New York; PCP primary care physician; SA spondyloarthritis; SD standard deviation; VA Veterans Administration.



**Table 2. Comparison of Components Among Clinical Algorithms for Diagnosis of Gout**

Criteria Components	Wallace 1977 (ARA/ ACR) <sup>9</sup>	Rome <sup>a27</sup>	NY <sup>b27</sup>	EULAR 2006 <sup>14</sup>	Janssens 2010 <sup>28</sup>	CGD 2010 <sup>c41</sup>	3e Initiative 2014 <sup>d42</sup>	Richette Survey <sup>e30</sup>
>1 attack of acute arthritis	✓				✓	✓		✓
Maximum inflammation developed within 1 day	✓	✓ painful joint swelling, abrupt onset, clearing 1-2 wk	✓ 2 attacks	✓	✓	✓		
Monoarthritis/oligoarthritis attack	✓					✓		
Redness observed over joints	✓			✓	✓	✓		
1 <sup>st</sup> MTP joint painful or swollen	✓			✓	✓	✓ podagra	✓ podagra	
Unilateral 1 <sup>st</sup> MTP joint attack	✓		✓					
Unilateral tarsal joint attack	✓					✓		
Tophi (proven or suspected)	✓	✓	✓	✓		✓	✓	✓
Hyperuricemia	✓	✓		✓	✓	✓		✓
Asymmetric swelling within a joint on radiograph	✓			✓				
Subcortical cysts without erosions on radiograph	✓			✓				
Joint fluid culture negative	✓							
MSU crystals in synovial fluid or tissues		✓					✓	
Abrupt onset and remission in 1-2 weeks initially			✓					
Response to colchicine – major reduction in inflammation within 48h			✓				✓	
Male sex					✓			✓
Hypertension or ≥1 CVD					✓			✓
Hypertriglyceridemia								✓
Pain intensity ≥9/10								✓
Involvement of toes, foot, or ankles								✓
Treatment with								✓

corticosteroids								
Treatment with NSAIDS								✓
Resolution of pain<15 days after onset								✓

Table Notes: <sup>a</sup>Meets 2 of the following criteria; <sup>b</sup>MSU crystals in joint fluid or tophus or tissue OR meets 2 of the following criteria; <sup>c</sup>≥4/8 of the following criteria; <sup>d</sup>Guideline 1 states that MSU is required for a definitive diagnosis but in its absence, clinical criteria such as the following can be used or characteristic imaging findings may substitute; <sup>e</sup>designed to be administered telephonically by non-physicians to assess prevalence of gout via patient self-report; treatment questions refer to most prominent episode.

## Studies Assessing DECT to Diagnose Gout

### Comparative Accuracy

We identified three original studies that compared the use of DECT for gout diagnosis to a reference standard and that enrolled only patients suspected of having gout.<sup>10, 32, 33</sup> We also identified one 2014 systematic review comparing DECT and US for the classification of gout.<sup>37</sup> The study details are described in Table 3.

A 2011 U.S. study used DECT to assess 94 patients with suspected gout. Of the 94, 31 subsequently underwent successful joint aspiration: 12 tested positive for uric acid crystals in the aspirated joint and 19 were negative. Two blinded radiologists read the images, finding a sensitivity of 100% and specificities of 89% and 79% ( $K=0.87$ ). On the basis of these findings, the authors concluded that DECT is a sensitive and reproducible method for uric acid detection in multiple anatomical sites, including several atypical sites for gout.<sup>10</sup>

A 2014 U.S. study to determine the diagnostic accuracy of DECT examined 81 patients with suspected gout, 40 of whom were diagnosed with acute gout, and 41 with other arthritic conditions, at a single center.<sup>32</sup> This assessment showed a sensitivity of 90% and a specificity of 83%. All false-negative DECT scans were patients with acute, recent-onset gout; all false positives had advanced knee osteoarthritis. In a separate study, they then studied 30 patients with synovial arthritis but an unclear diagnosis; 16 of the 30 showed signs of gout on DECT, and 14 were DECT positive for indications of gout. Among the DECT-positive patients, two refused subsequent joint aspiration, and 11 of the 12 who underwent aspiration were positive for MSU crystals for a specificity of 92%. The DECT-negative patients did not undergo aspiration, so sensitivity could not be determined for this group. The authors concluded that while DECT is diagnostically useful overall, it has limited sensitivity among patients with acute gout and no prior history of gouty arthritis.

A 2014 study conducted in Germany at a single center assessed 60 patients with suspected gout, 39 of whom were subsequently diagnosed with gout either through polarization microscopy or use of a validated clinical algorithm, and the diagnostic outcomes were compared with those of DECT and of US.<sup>33</sup> They found that DECT gave a sensitivity of 85% and a specificity of 86% (positive predictive value [PPV] 92%, negative predictive value [NPV] 75%). The comparison of DECT with US is described below. .

A 2014 systematic review (AMSTAR score: 6 of 11)<sup>37</sup> pooled the results of two of our include studies.<sup>10, 32</sup> And one additional study of confirmed gout patients.

### **The Effect of Test Accuracy on Clinical Decision Making, Clinical Outcomes and Complications, and Patient Centered Outcomes**

No studies were identified that tested the effects of the accuracy of gout diagnoses based on DECT on clinical decision making (treatment decisions or decisions to conduct further testing), clinical outcomes, or patient centered outcomes.

### **The Effect of Affected Joint Site(s) and Number of Joints on the Diagnostic Accuracy of DECT**

No studies were identified that tested the effects of the joint site affected or the number of joints affected on the accuracy of DECT.

### **The Effect of Duration of Symptoms on the Accuracy of Gout Diagnoses Based on DECT**

No studies tested the effect of duration of symptoms (time since beginning of flare) on the accuracy of gout diagnoses based on DECT.

**Table 3. Studies Assessing the Accuracy of DECT to Diagnose Gout**

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Test Components/Reference Standard	Outcome Measures and Findings	Conclusions and Other Comments
Huppertz 2014 <sup>33</sup> Germany 1 academic rheumatology department	60 consecutive patients with suspected gout (gout (n=39); remainder: rheumatoid arthritis (n=6); psoriatic arthritis (n=4); undifferentiated oligoarthritis (n=3); chondrocalcinosis (n=2); osteoarthritis (n=2); ankylosing spondylitis, undifferentiated tissue disease, Lyme disease, hydroxyapatite deposition disease (n=1, respectively) <b>Mean age:</b> 62 <b>%female:</b> 18%	Gout suspects  39 diagnosed with gout:  Polarization microscopy of synovial fluid (n=18/39 = 46%)  OR  Positive Janssens Score by rheumatologist (≥8 out of a maximum of 13 points) incorporating serum uric acid level (>350micromol/L; 3.5 points), first MTP joint involvement (2.5 points), gender (male, 2 points), previous patient-reported arthritis attack (2 points), cardiovascular diseases (1.5 points), joint redness and onset within 1 day (0.5 points) (n=21/39 = 54%)	Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value of DECT and US	<b>For gout diagnosis:</b> DECT Sensitivity: 84.6% DECT Specificity: 85.7%%  US Sensitivity: 100% US Specificity: 76.2%  DECT Pos. Predictive Val.: 91.7% DECT Neg. Predictive Val.: 75.0%  US Pos. Predictive Val.: 88.6% US Neg. Predictive Val.: 100%  All patients with false-positive DECT were negative in US; all patients with false- positive US were negative in DECT.  <b>Comparing 18 MSU+ positive gout patients and 21 patients with other rheumatic disease:</b>  DECT Sensitivity: 83.3% US Sensitivity: 100%  DECT Pos. Predictive Val.: 83.3% US Pos. Predictive Val.: 78.3%  DECT Neg. Predictive Val.: 85.7% US Neg. Predictive Val.: 100% Authors conclude that DECT provides better differential diagnosis than US  Data provided on detection of crystals by joint (KQ1b?)
Bongartz 2014 <sup>32</sup> US 1 academic	Accuracy study: 81 patients (40 with acute gout; 41 other)	Polarizing and electron microscopy of synovial fluid	DECT Sensitivity and Specificity Adverse events	<b>Accuracy study:</b> Interobserver Kappa for polarizing microscopy= 0.93

rheumatology department	<p>Diagnostic yield study: 30 patients with ambiguous findings (tendonitis but MSU- [22] or inadequate synovial fluid [8]) <b>Mean age:</b> 62.1 (gout); 58.7 (non-gout) <b>%female:</b> 9% (gout); 29% (non-gout)</p>			<p>Sensitivity = 90% Specificity = 83%</p> <p><b>Diagnostic yield study:</b> Patients with inflammatory arthritis + no synovial crystals n = 30</p> <ul style="list-style-type: none"> <li>- In this group, DECT negative n = 16</li> <li>- In this group, DECT positive n = 14</li> <li>- In DECT positive group, US guided aspiration positive n = 11/12</li> <li>- In DECT positive group, US guided aspiration negative n = 1/12 (since 2 declined aspiration)</li> <li>- Specificity=92%</li> </ul> <p>Study reports no adverse events from DECT of synovial fluid aspiration to analyze MSU crystals</p>
Glazebrook 2011 <sup>10</sup> US 1 academic or rheumatology department	<p>94 gout suspects, of whom 43 underwent attempted joint aspiration and 31 patients had successful aspiration (12 crystal-positive; 19 crystal-negative); <b>Mean age:</b> n/a <b>%female:</b> n/a</p>	Reference standard: Joint aspiration for uric acid crystals	Accuracy, Sensitivity, Specificity, and Interobserver agreement of DECT (two readers)	<p>Readers 1 and 2 had no false negative findings for uric acid through DECT.</p> <p>DECT Sensitivity = 100% DECT Specificity (Reader1) = 89% DECT Specificity (Reader2) = 79%</p> <p>Interobserver agreement K = 0.87</p>
Ogdie 2014 <sup>37</sup> Systematic review of gout diagnosis in primary care settings	<p>Inclusion criteria: studies comparing diagnostic performance of X-Ray, MRI, US, CT or DECT in gout; presence of non-gout control group; gout confirmation through synovial fluid aspiration.</p> <p>Exclusion criteria: use of clinical criteria or physician or patient reports</p>	Joint aspiration for uric acid crystals	Sensitivity and Specificity of: Double-Contour Sign (DCS) on Ultrasound; Ultrasound of Tophus; DECT.	<p>Ultrasound DCS Sensitivity (pooled) = 83% Ultrasound DCS specificity (pooled) = 76%</p> <p>Ultrasound tophus sensitivity (pooled) = 65% Ultrasound tophus specificity (pooled) = 80%</p> <p>DECT sensitivity (pooled) = 87% DECT specificity (pooled)</p>

	instead of MSU crystal presence; lack of control or comparison group; cases with asymptomatic hyperuricemia; not enough information to calculate sensitivity and specificity			
--	--	--	--	--

Table Notes: ACR American College of Rheumatology; ARA American Rheumatism Association; AUC area under the curve; BMI body mass index; CGD Clinical Gout Diagnosis; CI confidence interval; CRP c-reactive protein; CVD cardiovascular disease; GFR glomerular filtration rate; HTN hypertension; LR likelihood ratio; MSU monosodium urate; MTP metatarsophalangeal; OA osteoarthritis; NPV negative predictive value; NY New York; PCP primary care physician; SA spondyloarthritis; SD standard deviation; VA Veterans Administration.





## Studies Assessing the Accuracy of Ultrasound to Diagnose Gout

### Comparative Accuracy

We identified four original studies that assessed the accuracy of ultrasound (US) to diagnose gout in patients suspected of having gout compared to a reference standard comprising MSU crystal analysis, the use of a clinical algorithm, or a combination of the two.<sup>33-36</sup> These studies are described in Table 4. The QUADAS ratings for each of these studies are presented in Table 5. We also identified two recent systematic reviews on the use of US to diagnose gout.<sup>37, 38</sup>

A 2008 study by Rettenbacher conducted in Austria with a population of 86 patients with clinical suspicion of gout compared US with plain x-ray and with a reference standard of MSU assessment (30 patients) or a clinical algorithm (56 patients).<sup>35</sup> The authors assessed the sensitivity and specificity for a number of US indicators, included bright stippled foci (BSF), hyperechoic areas, at least one of the former two, bone erosions, and hyperechoic streaks. For US, BSF and hyperechoic areas had the highest specificity for gout (hyperechoic areas had a much higher specificity than BSF with comparable sensitivity). Compared with US, plain x-ray was less specific but much more sensitive.

A 2011 study by Lai that was conducted in a Taiwan hospital enrolled 80 patients with monoarthritis. US of the target joint was compared with joint aspiration and fluid analysis for MSU and a clinical algorithm.<sup>36</sup> Of the 80 patients, 34 (representing 52 joints) were MSU+. Images were assessed for double contour sign (DCS), snow storm, tophi, and bony erosion. The DCS alone had a PPV of 93 and a NPV of 60; the presence of DCS or BSF had a PPV of 70 and a NPV of 82; whereas the DCS combined with the BSF had a PPV of 100 and a NPV of 56. Interobserver correlations were excellent for the DCS (K=0.86) and good for BSF (0.73).

A 2014 study conducted in the Netherlands enrolled 54 patients with monoarthritis but no further diagnosis.<sup>34</sup> Patients underwent US of the knees, MTP joints, and any additional target joint of interest, bilaterally, for a total of 6 images; the images were examined for double contour sign, snow storm, tophi, and any US finding. Assessments were compared with the results of joint aspiration and MSU analysis or a clinical algorithm (not described in detail). The finding of a double contour sign had a sensitivity of 77% and a specificity of 75% inclusion of any sign had a sensitivity of 96% and snowstorm had a specificity of 86%).

A 2014 study conducted in Germany and described above compared the findings of US with those of DECT<sup>33</sup> among 60 gout suspects. The assessment of hyperechoic findings and DCS had a sensitivity of 100% and specificity of 76% (PPV 89, NPV 100). The comparison with DECT showed that DECT is more specific than US for the diagnosis of gout and is particularly useful for patients with non-conclusive joint aspiration, simultaneous rheumatic diseases, and ambivalent findings. In a joint-based evaluation, US was more sensitive than DECT for gout diagnosis; however, differentiating between CPPD and gout at the wrist is challenging with US.

Two systematic reviews assessed studies of US for gout. A 2012 systematic review by Chowaloor (AMSTAR score: 4 of 11)<sup>38</sup> reviewed the results of 18 studies (1975 to 2012) on the use of US for management (and secondarily, diagnosis) of gout. Only one of these studies was included in the present review;<sup>35</sup> the remaining studies were excluded either because they included only patients with definitive gout diagnoses, included only patients with chronic gout, or included primarily patients with asymptomatic hyperuricemia. The review, which did not include a meta-

analysis, concluded that the DCS is specific for the diagnosis of gout, and that US can track changes in joint architecture, e.g., development of tophi.

A 2014 systematic review (AMSTAR score: 6 of 11)<sup>37</sup> on the use of US (and DECT) for gout classification pooled the results of five studies, one of which was included in the present review<sup>34</sup> (the remaining four were excluded because the population consisted of patients with prior gout diagnoses). That review reported a sensitivity of 0.65, a specificity of 0.80, and an AUC of 0.75 for US in detecting gout.

### **The Effect of Test Accuracy on Clinical Decision Making, Clinical Outcomes and Complications, and Patient Centered Outcomes**

No studies were identified that tested the effects of the accuracy of gout diagnoses based on ultrasound on clinical decision making (treatment decisions or decisions to conduct further testing), clinical outcomes, or patient centered outcomes.

### **The Effect of Affected Joint Site(s) and Number of Joints on the Diagnostic Accuracy of DECT**

No studies were identified that tested the effects of the joint site affected or the number of joints affected on the accuracy of diagnosis by ultrasound.

### **The Effect of Duration of Symptoms on the Accuracy of Gout Diagnoses Based on Ultrasound**

No studies tested the effect of duration of symptoms (time since beginning of flare) on the accuracy of gout diagnoses based on ultrasound.

**Table 4. Summary of Studies Reporting on the Use of Ultrasound for Diagnosis of Gout**

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Criteria/ reference standard	Joints	US characteristics	Outcomes	Author Conclusions Comments
Lamers-Karnebeck, 2014 <sup>34</sup> Netherlands Academic Rheumatology Clinic	Monoarthritis (n=54) 28 gout dx (26 MSU+) 7 CPPD 19 non-crystal arthritis Mean age: MSU+: 63.5 (55.5–69.5) MSU-: 55.0 (41.8–63.5) % female: MSU+: 4% MSU-: 86.6	Aspiration Clinical Not well described	6 joints Target, Knee, MTP (bilateral)	Any US Finding Double Contour Snow Storm Tophi	<u>Any abnormality:</u> Sensitivity:96 Specificity:68 PPV:74 NPV:95 LR+:2.99 LR-:0.06 <u>Double contour sign:</u> Sensitivity:77 Specificity: 75 PPV: 74 NPV: 78 LR+: 3.08 LR-: 0.31 <u>Snow storm:</u> Sensitivity:38 Specificity: 86 PPV: 71 NPV: 60 LR+: 2.6 LR-: 0.72 <u>Tophus Presence:</u> Sensitivity: 19 Specificity: 93 PPV: 71 NPV: 55 LR+:2.69 LR-: 0.87	Authors conclude that US, if done by skilled diagnosticians has a useful place as part of early gout diagnostic algorithm. DCS is seen in 77% of gout patients and 25% of non-gout patients. Comment: Best for purpose of utility of diagnosis
Lai, 2011 <sup>36</sup> Taiwan Hospital rheumatology department	Monoarthritis (n=80) 34 (52 joints) MSU+ 46 (52 joints) MSU- Mean age: 69.4 (range 52.8-80.8) % female: 17.6% (gout) 52.2% (non-gout)	Aspiration Clinical criteria (joint swelling, heat, tenderness)	Target Joint	Double Contour Snow Storm Tophi Bony erosion PDUS	<u>Double contour sign (DCS):</u> Sensitivity: 36.8 Specificity: 97.3 PPV: 93.3 NPV: 60 LR+: 13.63 LR-: 0.65 <u>Bright Stippled foci (BSF):</u> Sensitivity: 76.9	Analysis was at the joint level, not patient Authors conclude that presence of both DCS and BSF is indicative of gout.

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Criteria/ reference standard	Joints	US characteristics	Outcomes	Author Conclusions Comments
					Specificity: 65.4 PPV: 69 NPV: 73.9 LR+: 2.22 LR-: 0.35 <u>DCS OR BSF:</u> Sensitivity: 86.5 Specificity: 63.5 PPV: 70.3 NPV: 82.5 LR+: 2.37 LR-: 0.21 <u>DCS AND BSF:</u> Sensitivity: 23.7 Specificity: 100 PPV: 100 NPV: 56.1 LR+: LR-: 0.76	Hyperechoic aggregations, bone erosion, and synovial vascularity did not distinguish gout from non-gout arthritis.
Huppertz, 2014 <sup>33</sup> Germany Academic rheumatology department	60 patients with suspected gout: Gout (39) RA (6) PsA(4) Other (8) CPPD (2) Hydroxyapatite disease (1) Mean age: 62[11.3] (range: 36-82) % female: 18.3	Clinical impression: Crystal +/- clinical score (Janssens≥8 considered alternate standard of reference for gout)	Foot/ankle Knee Hand/wrist Elbow	Hyperechoic findings Double contour sign	US: Sensitivity: 100 Specificity: 76.2 PPV: 88.6 NPV: 100 LR+: 4.20 LR-: 0  DECT: Sensitivity: 84.6 Specificity: 85.7 PPV: 91.7 NPV: 75 LR+: 5.92 LR-: 0.18	Study compares US with DECT: US has lower specificity than DECT for gout but is more sensitive to small MSU crystal deposits

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Criteria/ reference standard	Joints	US characteristics	Outcomes	Author Conclusions Comments
Rettenbacher, 2008 <sup>35</sup> Austria Academic radiology, rheumatology, and internal medicine departments	105 patients with clinical suspicion of gout Mean age 59 (range: 31-89) % female: 12% Gout(55) Another disease (31) No definitive diagnosis (19, excluded)	Aspiration of synovial fluid and MSU crystal analysis (30 patients) or characteristic clinical and lab findings (56 patients, not further described)	Multiple sites including MTP1, PIP joint (foot and hand), DIP joint, tarsus, ankle, knee, MCP joint, Metacarpus, Carpus, elbow, Achilles tendon, tibialis anterior tendon, ligamentum patellae, extensor tendons, tendons of third finger	Bright stippled foci Hyperechoic areas Bone erosions Hyperechoic streaks Hypervascularization	<p><b>US:</b> <b>BSF:</b> Sensitivity: 80 Specificity:75 PPV: 85 NPV: 69 LR+: 3.2 LR-:0.27</p> <p><u>Hyperechoic areas:</u> Sensitivity: 79 Specificity: 95 PPV: 96 NPV: 73 LR+:15.8 LR-:0.22:</p> <p><u>At least one of the above +:</u> Sensitivity: 96 Specificity: 73 PPV: 86 NPV: 91 LR+: 3.56 LR-:0.05</p> <p><u>Bone erosions:</u> Sensitivity: 24 Specificity:69 PPV: 57 NPV: 34 LR+:0.77 LR-:1.10</p> <p><u>Hyperechoic streaks:</u> Sensitivity: 80 Specificity:49 PPV: 1.57 NPV: 0.41</p> <p>Hypervascularization</p>	US compared with plain x- ray For US, alone, BSF and hyperechoic areas are most suggestive of gout. X-ray is less specific but much more sensitive for diagnosing gout than is US

Author, Year Country Practice setting	Population n, putative diagnosis, mean age, % female	Criteria/ reference standard	Joints	US characteristics	Outcomes	Author Conclusions Comments
					Sensitivity: 94 Specificity: 53 PPV: 77 NPV: 84 LR+: 2.00 LR-: 0.11  <b>X-ray:</b> <u>Opacification:</u> Sensitivity: 26 Specificity: 97 PPV: 93 NPV: 43 LR+: 8.67 LR-: 0.76  <u>Bone erosions:</u> Sensitivity: 20 Specificity: 95 PPV: 87 NPV: 41 LR+: 4.00 LR-: 0.84  <u>Osseous Apposition:</u> Sensitivity: 5 Specificity: 100 PPV: 100 NPV: 38 LR+: LR-: 0.95  <u>At least one positive sign:</u> Sensitivity: 31 Specificity: 93 PPV: 89 NPV: 44 LR+: 4.43 LR-: 0.74	

Table notes: AH= Asymptomatic Hyperuricemia; BSF= Bright Stippled foci; CPPD= Calcium Pyrophosphate Deposition (formerly pseudogout); DECT= Dual-energy computed tomography; DCS= Double-Contour Sign; LR= Likelihood ratio; MSU= Monosodium urate; MTP= Metatarsophalangeal; NPV= Negative predictive value; PPV= Positive predictive value; PsA= psoriatic arthritis; RA= Rheumatoid arthritis; US= Ultrasound





**Table 5. Risk of bias assessment by QUADAS-2**

Author, year	Domain 1: Patient Selection				Domain 2: Index Test(s)			Domain 3: Reference Standard			Domain 4: Flow and Timing				
	Consecutive/ random sample	Case-control design avoided?	Appropriate exclusions avoided?	Overall risk of bias	Index test results read without knowing reference standard?	Prespecified threshold?	Overall risk of bias	Reference standard correctly classifies target condition?	Reference standard read without knowing the index test?	Overall risk of bias	Appropriate interval between index test(s) and reference standard <sup>2</sup>	All patients received reference standard?	Same reference standard?	All patients included in the analysis?	Overall risk of bias
Bongartz et al., 2014 <sup>32</sup>	Yes	Yes	Unclear	<b>Low</b>	Yes	Yes	<b>Low</b>	Yes	Yes	<b>Low</b>	Yes	Yes	Yes	Yes	<b>Low</b>
Glazebrook et al., 2011 <sup>10</sup>	Yes	Yes	Yes	<b>Low</b>	Yes	N/A	<b>Low</b>	Yes	Unclear	<b>Unclear</b>	No	Yes	Yes	No	<b>Unclear</b>
Hasselbacher et al., 1987 <sup>43</sup>	N/A	N/A	No	<b>High</b>	N/A	N/A	<b>N/A</b>	Yes	N/A	<b>Low</b>	N/A	Yes	Yes	Yes	<b>Low</b>
Huppertz et al., 2014 <sup>33</sup>	Yes	No	Unclear	<b>High</b>	Yes	N/A	<b>Low</b>	Yes	Yes	<b>Low</b>	Yes	Yes	Yes	Yes	<b>Low</b>
Janssens, et al., 2010 <sup>29</sup>	Yes	Yes	Unclear	<b>Low</b>	Unclear	Yes	<b>Unclear</b>	Yes	Unclear	<b>Unclear</b>	Yes	Yes	Yes	Yes	<b>Low</b>
Janssens et al., 2010 <sup>28</sup>	Yes	Yes	Yes	<b>Low</b>	Yes	Yes	<b>Low</b>	Yes	Yes	<b>Low</b>	Yes	Yes	Yes	Yes	<b>Low</b>
Kienhorst et al., 2013 <sup>24</sup>	Unclear	Yes	Unclear	<b>Unclear</b>	Yes	Yes	<b>Low</b>	Yes	Yes	<b>Low</b>	Unclear	Yes	Yes	Yes	<b>Low</b>
Kienhorst et al., 2014 <sup>26</sup>	Yes	Yes	No	<b>Low</b>	Yes	N/A	<b>Low</b>	Yes	No	<b>Unclear</b>	Yes	Yes	Yes	Yes	<b>Low</b>
Lai et al., 2011 <sup>36</sup>	Yes	Yes	Unclear	<b>Low</b>	Yes	N/A	<b>Low</b>	Yes	Yes	<b>Low</b>	Unclear	Yes	Yes	Yes	<b>Low</b>
Lamers-Karnebeek et al., 2014 <sup>34</sup>	Yes	Yes	No	<b>Low</b>	Yes	N/A	<b>Low</b>	Yes	Yes	<b>Low</b>	Yes	Yes	Yes	Yes	<b>Low</b>
Lenski et al., 2014 <sup>25</sup>	Yes	No	Yes	<b>High</b>	Unclear	Yes	<b>Unclear</b>	Yes	Yes	<b>Low</b>	Yes	Yes	Yes	Yes	<b>Low</b>
Malik et al., 2009 <sup>27</sup>	Yes	Yes	Yes	<b>Low</b>	Yes	Yes	<b>Low</b>	Yes	Yes	<b>Low</b>	Unclear	Yes	Yes	Yes	<b>Low</b>
Park et al., 2014 <sup>44</sup>	Yes	Yes	No	<b>Low</b>	N/A	N/A	<b>N/A</b>	Yes	N/A	<b>Low</b>	N/A	Yes	Yes	No	<b>Unclear</b>
Rettenbacher et al., 2008 <sup>35</sup>	Yes	Yes	Yes	<b>Low</b>	Yes	N/A	<b>Low</b>	Yes	Unclear	<b>Unclear</b>	Unclear	Yes	No	No	<b>High</b>
Richette et	Unclear	No	Unclear	<b>High</b>	Yes	No	<b>Unclear</b>	Yes	Yes	<b>Low</b>	Unclear	Yes	Yes	Yes	<b>Unclear</b>

Author, year	Domain 1: Patient Selection				Domain 2: Index Test(s)			Domain 3: Reference Standard			Domain 4: Flow and Timing				
	Consecutive/ random sample	Case-control design avoided?	Appropriate exclusions avoided?	Overall risk of bias	Index test results read without knowing reference standard?	Prespecified threshold?	Overall risk of bias	Reference standard correctly classifies target condition?	Reference standard read without knowing the index test?	Overall risk of bias	Appropriate interval between index test(s) and reference standard?	All patients received reference standard?	Same reference standard?	All patients included in the analysis?	Overall risk of bias
al., 2014 <sup>30</sup>															
Vazquez-Mellado et al., 2012 <sup>31</sup>	Yes	No	Unclear	<b>High</b>	No	Yes	<b>High</b>	Yes	Yes	<b>Low</b>	Unclear	Yes	Yes	Yes	<b>Low</b>
Wallace et al., 1977 <sup>9</sup>	Unclear	No	Unclear	<b>High</b>	Unclear	Yes	<b>High</b>	Yes	Yes	<b>Low</b>	No	No	Yes	Unclear	<b>High</b>

Table notes: N/A= Not available

## KQ1d. Factors Affecting the Accuracy of Synovial Fluid Aspiration and Crystal Analysis

The question we were asked to answer was whether the accuracy of gout detection by synovial fluid aspiration and crystal analysis differ by i) the type of practitioner who is performing the aspiration and ii) the type of practitioner who is performing the crystal analysis.

### Key Points

- Detection of MSU crystals in synovial fluid appears to be affected by a number of factors, which include the experience and training of analysts, the storage and handling of samples, and patient factors such as the time lag between the onset of a gout attack and synovial fluid assessment. Because of the relatively small number of studies identified, **the strength of evidence for definitive influential factors is low.**

### Description of Included Studies

We identified one original study that addressed this question directly. We also identified one original study and one medium quality 2013 systematic review that addressed related issues that are also summarized here. A 2014 study was identified that retrospectively audited medical records of two Korean academic medical centers to assess factors associated with false negative synovial fluid MSU results and focused on the personnel performing the analysis and several other factors.<sup>44</sup> A 1987 survey examined the accuracy of MSU and CPPD analysis among hospitals and hospital laboratories in two states.<sup>43</sup> Finally, a 2014 systematic review examined the accuracy of methods for MSU crystal detection in synovial fluid samples and the effects of sample handling.<sup>39</sup>

### Detailed Synthesis

A 2014 study examined retrospective data from 179 patients (94.4% male, average age at diagnosis 62.6[16.4], 43.9% were experiencing their first attack) seen in two South Korean academic rheumatology departments to assess the rate of MSU detection and factors associated with false-negative findings.<sup>44</sup> The charts of patients who underwent synovial fluid analysis for presumed acute gout between 1999 and 2011 were audited; patients were excluded if they were diagnosed with intercritical gout without acute symptoms, pseudogout, OA, or both SA and gout. Patients were classified as having acute gout based on ACR criteria. MSU analysis at both hospitals was conducted either by lab medicine personnel (34.3%), by Rheumatology Department personnel (3.5%), or both (62.1%), at both hospitals. Of 198 samples analyzed for the 179 patients, 78.8% were crystal-positive, and 21.2% were negative. The detection rate for samples examined by the lab medicine departments was 51.8%, compared with 93.8% for the rheumatologists. When agreement was assessed for samples analyzed by both departments, the detection rate was 93.5% for rheumatologists and 51.2% for lab medicine, for a kappa of 0.108. Examination by lab medicine was the most significant variable associated with crystal-negative synovial fluid analysis (OR 36.996, 95% CI 9.731, 140.648). In addition, multivariate analysis showed that the time interval from attack onset to arthrocentesis was the only factor significantly associated with the likelihood of negative crystal findings (4.5±5.1 days for negative findings vs. 3.0±2.8 days for positive findings; OR 1.105, 95% CI 1.004, 1.216; p=0.042). Analysis by

laboratory medicine was thought to be associated with a longer lag time than analysis by the Rheumatology departments.

A 1987 survey was conducted to assess the frequency and accuracy of synovial fluid analysis for CPPD and MSU crystals by all hospitals in the states of New Hampshire and Vermont.<sup>43</sup> Of 42 hospitals contacted, 39 responded and completed the survey, 26 agreed to conduct testing of blinded samples, and 25 reported their results. Of the hospitals surveyed, three had rheumatology house staff; these hospitals reported performing more analyses (corrected for bed number). The testing in all hospitals that responded was performed by lab personnel. The survey did not assess (or did not report) what personnel aspirated the fluid at their hospitals. A correct assessment of MSU was made by 20 of 25 hospitals, compared with 24 of 25 for no crystals, only 4 of 25 for CPPD, and 1 of 25 for both CPPD and MSU. Thus, the accuracy of MSU assessment actually exceeded that for CPPD, a finding the authors attributed to the high concentration of MSU crystals in the MSU-only test solutions. Survey respondents asserted that training in synovial fluid analysis needed improvement.

A 2013 systematic review (AMSTAR 8/11) sought to answer the questions of the comparative accuracy of different methods of MSU crystal detection in synovial fluid compared with polarized light microscopy (PLM) and the effect of storage and handling factors on MSU crystal detection.<sup>39</sup> The review included 7 studies on comparative accuracy and 5 studies on sample storage and handling. Concordance with PLM of unstained samples was reported for PLM of Gram stained samples, Raman spectroscopy, Diff-Quick staining, Uricotest, electron microscopy, and ordinary light microscopy to detect MSU crystals, but risk of bias prevented firm conclusions. Studies on sample handling and storage suggested that these factors can affect analytic reliability: Storage at room temperature reduced MSU concentration compared with refrigeration, and dried cytopsin preparation increased stability compared with anticoagulation.

Key Question 2: What are the adverse effects associated with each diagnostic test (including pain, infection at the aspiration site, radiation exposure) or harms (related to false positives, false negatives, indeterminate results) associated with tests used to diagnose gout?

## Key Points

- No studies documented any adverse events associated with diagnostic tests in any studies included in this report. **The strength of evidence for this conclusion is low**, based on only one study that reported no adverse events associated with joint fluid aspiration for MSU analysis or DECT and no studies that reported on adverse events associated with US or clinical examinations.
- Misdiagnosis or delayed diagnosis of acute gout can result in unnecessary surgery, hospitalization, and delays in adequate treatment, although only one study was identified that examined outpatients with gout.
- **The strength of evidence for this conclusion is low.** The conclusion is based on one retrospective study in a non-U.S. hospital, but the findings of this study strongly argue for additional studies to be conducted.

## Description of included studies

One study was identified that assessed adverse effects associated with tests used to diagnose gout.<sup>32</sup> This study reported no adverse events associated with aspiration of synovial fluid for MSU analysis or the use of DECT.

One study examined the outcomes of delayed diagnosis or misdiagnosis of gout in two academic medical centers in South Korea.<sup>44</sup>

## Detailed Synthesis

A study that conducted retrospective medical chart review to examine factors associated with failure to diagnose gout also compared treatment outcomes between those who had a positive finding for MSU crystals and those who did not.<sup>44</sup> This study, described above, examined retrospective data from 179 patients (94.4% male, average age at diagnosis 62.6[16.4], 43.9% were experiencing their first attack) seen in two South Korean academic rheumatology departments. They found no differences in the proportions of MSU-negative and MSU-positive patients who received NSAIDs or colchicine, but MSU-negative patients were less likely than MSU-positive patients to receive intra-articular glucocorticoid injections (7.1% vs. 21.8%,  $p=0.043$ ). In addition, MSU-negative patients underwent emergency arthroscopic surgery and drainage more frequently than MSU-positive patients (26.2% vs. 3.8%,  $p=4.48 \times 10^{-6}$ , chi-squared test), based on a suspicion of septic arthritis, in spite of negative cultures and visible MSU crystals in the synovial fluid. A first episode of monoarthritis (OR 6.954, 95% CI 1.577, 20622,  $p=0.01$ ), negative findings for MSU crystal analysis (OR 23.760, 95% CI 4.451, 126.843  $p=2.10 \times 10^{-4}$ ), and fever (OR 15.123, 95% CI 2.739, 83.503,  $p=0.002$ ) were independent risk factors for surgery. Patients undergoing arthroscopic surgery had a significantly longer hospital stay than the 19 newly diagnosed gout patients hospitalized for medical treatment ( $18.4 \pm 12.6$  days vs.  $8.9 \pm 6.7$  days,  $p=0.007$ ), and patients undergoing arthroscopic surgery received anti-inflammatory drugs

significantly later than those who received medical treatment ( $5.0 \pm 3.4$  days vs.  $0.4 \pm 0.8$  days,  $p=8.74 \times 10^{-5}$ ).

# Discussion

## Key Findings and Strength of Evidence

The key findings and strength of evidence are summarized below and in Table 6.

### Accuracy of Tests for the Diagnosis of Gout

- Few studies that assessed the accuracy of clinical signs and symptoms consistently applied the same reference standard to all participants with suspected gout and sometimes included participants with chronic gout.
- Studies that assessed the use of clinical algorithms reported widely varying sensitivities and specificities; however, an algorithm developed from clinical signs and symptoms used by primary care physicians reported good positive and negative predictive value and was validated in a small secondary care population but needs further validation. **The strength of evidence for this conclusion is low, based on the identification of only two studies that assessed this particular clinical algorithm.**
- In small numbers of studies that enrolled only patients not previously diagnosed with gout, both ultrasound and DECT had good sensitivity and specificity for predicting gout, compared with synovial fluid analysis for MSU crystals or a validated clinical algorithm. Three studies of DECT revealed sensitivities that ranged from 85% to 100% and specificities that ranged from 83% to 92%. Four studies of US showed sensitivities that ranged from 37% to 100% and specificities that ranged from 68% to 97%, depending on the signs assessed. **The strength of evidence for this conclusion is low.**
- No studies were identified that assessed the clinical utility of serum uric acid, CT scan, or plain x-ray for diagnosing gout. **The strength of evidence for these tests is insufficient.**
- No studies were identified that directly assessed the effect of joint site or number of affected joints on diagnostic accuracy. **The strength of evidence for this question is insufficient for all diagnostic methods.**
- No studies were identified that directly assessed the effect of duration of symptoms on the accuracy of diagnostic tests. **The strength of evidence for this question is insufficient for all diagnostic methods.**

### Factors that Affect Accuracy of Synovial Fluid Analysis and MSU Crystal Analysis

- Detection of MSU crystals in synovial fluid appears to be affected by a number of factors, which include the experience and training of analysts, the storage and handling of samples, and patient factors such as the time lag between the onset of a gout attack and synovial fluid assessment. Because of the relatively small number of studies identified, **the strength of evidence for definitive influential factors is low.**

### Adverse Events Associated with Testing for Gout

- No studies documented any adverse events associated with diagnostic tests in any studies included in this report. **The strength of evidence for this conclusion is low, based on one study that reported no adverse events associated with joint fluid aspiration for MSU**

**analysis or DECT, and no studies that reported on adverse events associated with US or clinical examination.**

- Misdiagnosis or delayed diagnosis of acute gout can result in unnecessary surgery or hospitalization and delays in adequate treatment, although only one study was identified that examined outpatients with gout. **The strength of evidence for this conclusion is low.** The conclusion is based on one retrospective study in a non-U.S. hospital, but the findings of this study strongly argue for additional studies to be conducted

## **Findings in Relationship to What is Already Known**

### **Accuracy of Algorithms Comprising Clinical Signs and Symptoms for the Diagnosis of Gout**

Over the past 25 to 30 years, gout diagnosis has been an area of intense controversy. In 1989, Lichtenstein and Pincus, in an exchange of letters to the editor with Wallace regarding the validation of the ACR criteria, point out that the patients used to validate the criteria were handpicked by the 38 participating rheumatologists and had had gout for an average of 11 years and thus had more severe, well-established disease,<sup>24</sup> which would increase the sensitivity of the test compared with its use on the more typical population with less certain diagnosis.<sup>3</sup>

In 1994, 15 years before Malik compared the Rome, NY, and ACR criteria, Rigby and Wood<sup>45</sup> compared the Rome and NY criteria in a European population of patients diagnosed with gout by their own physicians and those without gout. The diagnostic criteria used to diagnose the patients with gout depended on the individual physician opinion. This study concluded that sUA was not a valid criterion for diagnosing gout, as there was no lower level below which gout was not a possibility. On this basis, they advocated its removal from the NY criteria. They also stated that response to colchicine was not tested sufficiently often to be considered a diagnostic criterion.

In 1999, Segal and Albert conducted a systematic review on the accuracy of MSU crystal analysis in synovial fluid.<sup>46</sup> The review, which included four studies, concluded that MSU analysis had poor sensitivity, specificity, and reproducibility. Its greatest use was thought to be to confirm the diagnosis in patients for whom clinical signs and symptoms had already indicated a high probability of disease.

Nevertheless, the 2006 EULAR diagnostic criteria for gout,<sup>14</sup> based on a literature review and expert panel consensus asserted that acute, short-lived pain, swelling may indicate gout, but that clinical diagnosis was not definitive without identification of MSU in synovial fluid or tophus. Synovial fluid MSU analysis was recommended in patients with no prior diagnosis of gout and in asymptomatic patients during intercritical periods (based on a study by Chen and Schumacher<sup>47</sup>). They also noted that because gout and septic arthritis can co-exist in the same joint, synovial fluid should be cultured if septic arthritis is suspected. The criteria also reaffirmed that serum UA cannot confirm or exclude a diagnosis of gout, that plain radiographs are not useful for diagnosis of early or acute gout, and that assessment of risk factors for gout can assist in diagnosis. Implementation of these guidelines has clearly been challenged, as evidenced by further research and recent efforts to set new guidelines for diagnosis and classification.



For example, a 2010 systematic review on the diagnosis of gout in women noted that clinical features and risk factors of gout in women differ from those in men (AMSTAR 6 of 11).<sup>40</sup> Women have later onset, are more likely to be taking diuretics, have more CVD and renal comorbidity, are less likely to drink alcohol, are less likely to have podagra (more involvement of other joints), are more likely to have polyarticular gout, and have less frequent recurrent attacks.

As we reviewed, in 2010, Janssens and colleagues conducted a series of prospective studies aimed at understanding and improving diagnostic criteria for gout in primary care settings.<sup>28</sup> The participants had suspected but not confirmed gout, and all underwent MSU crystal analysis. These diagnostic criteria have subsequently been validated in both primary<sup>29</sup> and secondary care populations.<sup>24</sup> The Diagnostic Rule, as it has come to be known, performed better than the ACR criteria in a comparable population.<sup>29</sup>

The 2011 Postgraduate Medicine guidelines for diagnosis of gout (which aimed to update the EULAR 2006 guidelines) emphasize reliance on clinical signs and symptoms, while acknowledging that MSU constitutes the definitive diagnosis.<sup>48</sup> Signs and symptoms include acute monoarticular attacks of the lower extremities; rapid development of severe pain, swelling, and tenderness that peaks within 6 to 12 hours, especially with overlying erythema is thought to be highly indicative of crystal formation, but not necessarily MSU crystals. The guidelines also note that diagnosis based on clinical signs and symptoms alone has reasonable accuracy when patients have typical presentation of gout, and that sUA levels cannot be used to diagnose or rule out gout. They also confirm the 2006 EULAR recommendation regarding performing Gram stain and culture if sepsis is suspected, as gout and septic arthritis can co-occur in the same joint. Finally, they assert that in a previous study, the presence of proven or suspected tophus and response to colchicine had the highest clinical diagnostic value (likelihood ratio) (although response to colchicine cannot distinguish gout from other crystal arthropathies and tophi are often not present in patients with recent onset gout).<sup>27</sup>

In a 2013 commentary, Dalbeth,<sup>49</sup> noting that none of the current diagnostic criteria have been adequately validated, describes an international project underway to establish and validate new gout classification criteria. The primary intent of this effort is to improve case ascertainment for recruitment into research studies and for epidemiological purposes. The need for this effort was predicted by the findings of a 2007 study by Harrold and colleagues that examined the validity of administrative data to estimate the prevalence of gout by comparing the diagnostic elements found in patient charts to the ACR, NY, and Rome criteria. A random sample of 200 patient charts was selected out of over 3000 with an ICD-9 code for gout in 4 U.S. HMOs, and pairs of rheumatologists abstracted diagnostic gout indicators. Indicators of gout in charts agreed poorly with the established criteria and with physician ratings, and inter-rater agreement was poor.<sup>50</sup> No studies that we identified assessed the time elapsed since the onset of the current attack, and few characterized the mean duration of a gout diagnosis among participants.

In 2014, the 3e (Evidence, Expertise, Exchange) initiative (3ei), a multinational effort to promote evidence-based practice, issued a set of recommendations on the diagnosis and treatment of gout, combining a literature review and expert opinion.<sup>42</sup> The 3ei recommendation on gout diagnosis recognizes the use of MSU as the gold standard but also the difficulty in performing this test under some circumstances, asserting that if MSU cannot be performed, the diagnosis “can be supported by classical clinical features (such as podagra, tophi, rapid response to colchicine), and/or characteristic imaging findings.” They note the poor diagnostic utility of most of the clinical criteria, by themselves, except response to colchicine and the presence of tophi; however the former cannot definitely distinguish gout from other crystal arthritis forms (and the latter may

be seen only in more advanced gout). In noting the potential superiority of newer imaging techniques, notably US and DECT, they acknowledge that “...availability, cost, and the need for trained personnel and specific equipment...” might limit their use in routine clinical practice. Thus, these guidelines seem to suggest that in primary care settings, diagnosis can be based on a set of clinical criteria.

### **Accuracy of DECT for the Diagnosis of Gout**

DECT is a non-invasive study method that can detect urate deposits in joints, tendons, bursa, and soft tissues. The radiographic signature of urate can be distinguished from that of calcium. DECT requires special machines and software to process the images and currently is not widely available. Radiation exposure is not greater than standard CT scanning and is limited to extremities, which are not radio-sensitive organs.

Studies looking at diagnostic utility of DECT are promising, demonstrating good sensitivity and specificity for gout. Bongartz<sup>32</sup> sought to determine the additive value of DECT to a clinically unclear presentation among 30 patients without clear diagnoses. Of these 30, 14 had a positive DECT, and 11 of 12 of those with positive DECT findings (2 patients refused aspiration) had crystal confirmation of gout using ultrasound guided aspiration, suggesting DECT may be a useful adjunct to clinical algorithms. However, among another group of 40 patients with newly diagnosed gout, all four patients with false negative DECT had new onset gout (first attack and symptom duration less than 6 months), suggesting DECT may be less useful in very early cases than in patients with disease of longer duration. Glazebrook had also prospectively studied inflammatory mono-arthritis patients, demonstrating high sensitivity and specificity for crystal-confirmed gout cases.<sup>10</sup>

The summary of the literature demonstrates that DECT is both specific and sensitive for gout. Utility of DECT may be best for evaluating urate burden in established gout patients. Limited data suggest that for patients with recurrent attacks of inflammatory mono- or oligo- arthritis where the question of gout is unresolved (for example, no fluid available for aspiration or negative study), DECT should demonstrate good diagnostic value. However, for patients with a first inflammatory mono-articular attack (due to gout), DECT may not be sensitive. The availability of DECT machines in most regions also may limit application of this technology.

### **Accuracy of Ultrasound for the Diagnosis of Gout**

Use of ultrasound as a diagnostic test for gout has promising potential. It is relatively inexpensive, non-invasive and well accepted by patients. Sensitivity and specificity for specific findings or amalgamation of findings were typically high.

However, several challenges must be overcome prior to US being accepted as a standard diagnostic technique for gout. Gout has several US characteristics, including the “double contour sign,” characteristic intra-articular findings (bright spots or “snow”), or tophaceous findings. As follows the nature of the disease, these findings can present in many different joints, and the analyses we reviewed each used different methodology for identifying which joints were studied. The number of joints studied ranged from a single target (inflamed) joint up to 26 joints. Additionally, up to 20 tendon areas and 6 bursae were also examined. Exhaustive scanning is not practical. Some authors (notably Lamers-Karnebeck)<sup>34</sup> described limited systematic evaluation of inflammatory mono-arthritis patients with sensitivities and likelihood ratios for specific findings. Nevertheless, this focused methodology (4 to 6 joints) is beyond what would be available from

most radiology centers, which typically focus on more comprehensive examinations of single joints. With most of these publications in rheumatology journals or by rheumatology authors, diagnostic enthusiasm for gout and such studies appears to be greatest in the rheumatology community. Furthermore, we did not find studies that evaluated the marginal utility of US data beyond clinical criteria or in lieu of joint aspiration.

## **Applicability**

Two factors may reduce the applicability of this review:

1. In many of the studies that assessed diagnostic algorithms, participants had already had a definitive diagnosis of gout, and sometimes had chronic gout. Relatively few of these studies enrolled only participants with suspected gout or monoarthritis. Three of the five studies included in the pooled analysis of the accuracy of DECT and three of the four studies included in the pooled analysis of US enrolled only gout suspects.
2. All studies were conducted in a rheumatology setting, usually an academic rheumatology department. Patients seen in this setting may have more advanced disease than those seen in a primary care setting, or may have comorbidities that add complexity to their treatment. A small number of studies purposefully recruited patients with PCP-diagnosed or suspected gout, expressly to assess the validity of PCP clinical decision making in gout diagnosis.

## **Implications for Clinical and Policy Decisionmaking**

The findings of this review provide some evidence to support the further development and validation of diagnostic algorithms based on a combination of clinical signs and symptoms for the diagnosis of gout in the primary care setting, with the use of imaging modalities (US and DECT) in cases where a definitive diagnosis cannot be made from signs and symptoms alone.

## **Limitations of the Comparative Effectiveness Review Process**

Assessing the comparative validity of diagnostic tests in systematic reviews presents a number of challenges that are not faced with comparative effectiveness reviews of treatment strategies. These limitations are magnified by several issues surrounding tests for gout and the natural history of the disease itself. To increase applicability, we limited included studies to those that enrolled previously undiagnosed patients; in doing so, we excluded a number of studies on the use of ultrasound and DECT for monitoring patients with chronic gout. Previous systematic reviews on the use of US and DECT included studies that enrolled patients with asymptomatic hyperuricemia and studies of patients with definitive gout diagnoses in various stages of the disease, as well as studies of patients with suspected gout (their findings were similar to ours).

In addition, our consideration of unpublished literature was limited. Although the Scientific Resource Center requested information from manufacturers of microscopes and imaging equipment used to diagnose gout (directly and through a notice in the Federal Register), no information was provided; we did not search FDA databases for such information ourselves. In addition, we included only conference proceedings cited in other included studies or suggested by TEP members or the subject matter expert on our project team.

## Limitations of the Evidence Base

The literature that addresses the diagnosis of gout has numerous limitations that make it difficult to draw firm conclusions. These limitations can be divided into three categories: study volume, design, and reporting quality.

### Volume

In short, few studies have attempted to address the diagnosis of gout. Almost no studies have examined the impact of diagnostic test accuracy on decision-making (decisions to order further testing or to initiate particular treatments) or any clinical or patient centered outcomes, and almost no studies addressed adverse events potentially associated with diagnostic testing. Most studies of gout address management issues or monitoring of patients with chronic gout.

### Design

Of the diagnostic studies we did identify, few studies limited enrollment to gout suspects or patients with a monoarthritis or some other clinical signs or symptoms that might suggest gout. Many studies enrolled patients with known gout, and many included no control group.

Even studies that enrolled patients who were gout suspects or included a control group systematically failed to limit enrollment to patients in their first attack or with recent onset or did not stratify findings by duration of the condition (as would be ascertained by asking, “How long have you been having these attacks?”). The lack of stratification by duration of condition would likely affect the positive and negative predictive value of imaging techniques more than it would affect diagnostic tests based on clinical signs and symptoms, but not necessarily, as one criterion in the latter is almost always the presence of tophus.

Most studies also fail to stratify by other relevant factors, such as time since the onset of the current or most recent flare, sex, and comorbidities. The time since onset of the current flare definitely affects the presence of crystals as well as clinical signs and symptoms.

No studies tested the validity of a diagnostic algorithm comprising clinical signs and symptoms and an imaging test, compared with clinical signs and symptoms or imaging alone.

Finally, issues concerning the use of synovial fluid MSU crystal identification as the reference standard abound. Taking the validity of the reference standard at face value, some studies assessed MSU in a fraction of participants only (e.g., those for whom synovial fluid could be aspirated, those most suspected of having gout, or those willing to undergo the test), using the ACR criteria or individual clinical judgment as the reference standard for the remaining participants. The technical problems with aspiration and analysis have been assessed and described extensively and include inconsistencies introduced by patient factors (e.g., the time lapse from the start of the flare to aspiration), sample handling factors (storage duration and temperature), and practitioner skills in aspiration and analysis.

### Reporting Quality

Failure to report important study design details in publications is a further limitation. Studies tended to be vague regarding blinding of assessors and the time lapse between implementation of the index test and reference standard (and the sequence of tests), a critical detail considering the short duration of gout attacks.

## Research Gaps

As described in the section on the limitations of the research base, above, promising algorithms have been validated in an insufficient number of studies, particularly studies in primary care settings.

Patient-level factors that influence test behavior have also been understudied: These include the influence of duration of a flare; number and identity of joints involved; and patient age, sex, and comorbidities.

Finally, studies may be needed to emphasize the impact of misdiagnosis of gout, either failure to diagnose gout and misdiagnosis as septic or osteoarthritis or failure to diagnose conditions such as septic arthritis.

## Conclusions

This review highlights the need to validate promising diagnostic algorithms in primary care settings, where the majority of patients with signs and symptoms suggestive of gout, but no definitive gout diagnosis, are likely to be seen. An algorithm with high diagnostic accuracy can ideally form part of a decision tree that combines clinical signs and symptoms with, or refers patients to rheumatologists for, more invasive tests or imaging for clinically ambiguous cases.

**Table 6. Summary of Findings and Strength of Evidence**

Key question	Number/type of studies	Strength of evidence	Findings
<b>1a Diagnostic accuracy</b>			
Clinical signs and symptoms (algorithms)	9 observational studies	Low	Tests vary in accuracy compared to synovial fluid aspiration and MSU crystal analysis. One algorithm based on primary care patients had AUC of 0.85 (95% CI 0.81-0.90)
Dual Emission Computerized Tomography (DECT)	3 observational 1 systematic review	Low	Sensitivity and specificity are good in patients with suspected gout
Ultrasound (US)	4 observational 2 systematic reviews	Low	Sensitivity and specificity are good in patients with suspected gout
Other tests	0 studies	Insufficient	None
<b>1b Influence of number and types of joints involved</b>	0 studies	Insufficient	None
<b>1c Influence of Symptom Duration</b>	0 studies	Insufficient	None
<b>1d. Influence of factors on analysis of monosodium urate crystals (MSU)</b>	2 observational studies 1 systematic review	Low	Results of MSU analysis appear to be affected by a number of factors related to patients (e.g., delayed presentation and joint aspiration), analyst experience, sample handling, and assay.
<b>2. Adverse events and implications of misdiagnosis</b>	2 observational studies: 1 on AEs associated with two diagnostic methods and 1 on implications of misdiagnosis	Low	One study reported DECT and joint aspiration for MSU analysis were associated with no adverse events. One study reported gout misdiagnosis resulted in longer hospital stays,

			unnecessary surgery, and delayed pharmacological treatment.
--	--	--	---

## References

1. Li C, Martin BC, Cummins DF, et al. Ambulatory resource utilization and cost for gout in United States. *American Journal of Pharmacy Benefits*. 2013 March/April;5(2):e46-e54. PMID: 2013264278.
2. Doghramji PP, Wortmann RL. Hyperuricemia and gout: new concepts in diagnosis and management. *Postgrad Med*. 2012 Nov;124(6):98-109. PMID: 23322143.
3. Wallace SL. Hyperuricemia in the diagnosis of gout. *J Gen Intern Med*. 1989 Mar-Apr;4(2):178-9. PMID: 2651605.
4. Singh JA, Reddy SG, Kundukulam J. Risk factors for gout and prevention: a systematic review of the literature. *Curr Opin Rheumatol*. 2011 Mar;23(2):192-202. PMID: 21285714.
5. Hamburger M, Baraf HS, Adamson TC, 3rd, et al. 2011 Recommendations for the diagnosis and management of gout and hyperuricemia. *Postgrad Med*. 2011 Nov;123(6 Suppl 1):3-36. PMID: 22156509.
6. Harrold LR, Mazor KM, Negron A, et al. Primary care providers' knowledge, beliefs and treatment practices for gout: results of a physician questionnaire. *Rheumatology (Oxford)*. 2013 Sep;52(9):1623-9. PMID: 23620554.
7. Owens D, Whelan B, McCarthy G. A survey of the management of gout in primary care. *Irish Medical Journal*. 2008;101(5).
8. Perez-Ruiz F, Carmona L, Yébenes MJG, et al. An Audit of the Variability of Diagnosis and Management of Gout in the Rheumatology Setting The Gout Evaluation and Management Study. *Jcr-Journal of Clinical Rheumatology*. 2011 Oct;17(7):349-55. PMID: WOS:000295320000001.
9. Wallace SL, Robinson H, Masi AT, et al. Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum*. 1977 Apr;20(3):895-900. PMID: 856219.
10. Glazebrook KN, Guimaraes LS, Murthy NS, et al. Identification of intraarticular and periarticular uric acid crystals with dual-energy CT: initial evaluation. *Radiology*. 2011 Nov;261(2):516-24. PMID: 21926378.
11. Khanna D, Fitzgerald JD, Khanna PP, et al. 2012 American college of rheumatology guidelines for management of gout. part 1: Systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. *Arthritis Care and Research*. 2012;64(10):1431-46.
12. Khanna D, Khanna PP, Fitzgerald JD, et al. 2012 American college of rheumatology guidelines for management of gout. part 2: Therapy and antiinflammatory prophylaxis of acute gouty arthritis. *Arthritis Care and Research*. 2012;64(10):1447-61.
13. Whiting P, Westwood M, Beynon R, et al. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol*. 2011 Jun;64(6):602-7. PMID: 21075596.
14. Zhang W, Doherty M, Pascual E, et al. EULAR evidence based recommendations for gout. Part I: Diagnosis. Report of a task force of the Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis*. 2006 Oct;65(10):1301-11. PMID: 16707533.

15. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011 Oct 18;155(8):529-36. PMID: 22007046.
16. Santaguida PL, Riley CM, Matchar DB. Assessing Risk of Bias as a Domain of Quality in Medical Test Studies. In: Chang SM, Matchar DB, Smetana GW, Umscheid CA, eds. *Methods Guide for Medical Test Reviews*. Rockville (MD); 2012.
17. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10. PMID: 17302989.
18. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001 Oct 15;20(19):2865-84. PMID: 11568945.
19. Team RDC. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>. 2014.
20. Chang SM, Matchar DB, Smetana GW, et al. *Methods Guide for Medical Test Reviews*. AHRQ Publication No. 12-EC017 Agency for Healthcare Research and Quality. Rockville (MD): 2012. <http://www.ncbi.nlm.nih.gov/pubmed/22834019>.
21. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994 Dec;50(4):1088-101. PMID: 7786990.
22. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997 Sep 13;315(7109):629-34. PMID: 9310563.
23. Atkins D, Chang S, Gartlehner G, et al. *Assessing the Applicability of Studies When Comparing Medical Interventions*. Agency for Healthcare Research and Quality; December 2010. *Methods Guide for Comparative Effectiveness Reviews*. AHRQ Publication No. 11-EHC019-EF.
24. Kienhorst L, Janssens H, Fransen J, et al. The performance of a diagnostic rule for acute gouty arthritis without joint fluid analysis: A validation study. *Arthritis and Rheumatism*. 2013 October;65 SUPPL. 10:S500.
25. Lenski M, Scherer MA. Analysis of synovial inflammatory markers to differ infectious from gouty arthritis. *Clin Biochem*. 2014 Jan;47(1-2):49-55. PMID: 24177196.
26. Kienhorst LB, Janssens HJ, Fransen J, et al. Arthritis of the first metatarsophalangeal joint is not always gout: A prospective cohort study in primary care patients. *Joint Bone Spine*. 2014 Jan 24; PMID: 24468668.
27. Malik A, Schumacher HR, Dinnella JE, et al. Clinical diagnostic criteria for gout: comparison with the gold standard of synovial fluid crystal analysis. *J Clin Rheumatol*. 2009 Feb;15(1):22-4. PMID: 19125136.
28. Janssens HJEM, Fransen J, Van De Lisdonk EH, et al. A diagnostic rule for acute gouty arthritis in primary care without joint fluid analysis. *Archives of Internal Medicine*. 2010;170(13):1120-6.
29. Janssens HJEM, Janssen M, Van De Lisdonk EH, et al. Limited validity of the American college of rheumatology criteria



for classifying patients with gout in primary care. *Annals of the Rheumatic Diseases*. 2010;69(6):1255-6.

30. Richette P, Clerson P, Bouee S, et al. Identification of patients with gout: Elaboration of a questionnaire for epidemiological studies. *Annals of the Rheumatic Diseases*. 2014.

31. Vazquez-Mellado J, Hernandez-Cuevas CB, Alvarez-Hernandez E, et al. The diagnostic value of the proposal for clinical gout diagnosis (CGD). *Clin Rheumatol*. 2012 Mar;31(3):429-34. PMID: 21979446.

32. Bongartz T, Glazebrook KN, Kavros SJ, et al. Dual-energy CT for the diagnosis of gout: an accuracy and diagnostic yield study. *Ann Rheum Dis*. 2014 Mar 25PMID: 24671771.

33. Huppertz A, Hermann KG, Diekhoff T, et al. Systemic staging for urate crystal deposits with dual-energy CT and ultrasound in patients with suspected gout. *Rheumatol Int*. 2014 Mar 12PMID: 24619560.

34. Lamers-Karnebeek FB, Van Riel PL, Jansen TL. Additive value for ultrasonographic signal in a screening algorithm for patients presenting with acute mono-/oligoarthritis in whom gout is suspected. *Clin Rheumatol*. 2014 Apr;33(4):555-9. PMID: 24510062.

35. Rettenbacher T, Ennemoser S, Weirich H, et al. Diagnostic imaging of gout: comparison of high-resolution US versus conventional X-ray. *Eur Radiol*. 2008 Mar;18(3):621-30. PMID: 17994238.

36. Lai KL, Chiu YM. Role of Ultrasonography in Diagnosing Gouty Arthritis. *Journal of Medical Ultrasound*. 2011 March;19(1):7-13. PMID: 2011227256

#### FULL TEXT LINK

<http://dx.doi.org/10.1016/j.jmu.2011.01.003>.

37. Ogdie A, Taylor W, Weatherall M, et al. Imaging modalities for the classification of gout: Systematic literature review and meta-analysis. *Annals of the Rheumatic Diseases*. 2014;73.

38. Chowalloor PV, Keen HI. A systematic review of ultrasonography in gout and asymptomatic hyperuricaemia. *Ann Rheum Dis*. 2013 May;72(5):638-45. PMID: 23291387.

39. Graf SW, Buchbinder R, Zochling J, et al. The accuracy of methods for urate crystal detection in synovial fluid and the effect of sample handling: a systematic review. *Clin Rheumatol*. 2013 Feb;32(2):225-32. PMID: 23138881.

40. Dirken-Heukensfeldt K, Teunissen TAM, van de Lisdonk EH, et al. "Clinical features of women with gout arthritis." A systematic review. *Clinical Rheumatology*. 2010 Jun;29(6):575-82. PMID: WOS:000277096500001.

41. Pelaez-Ballesteros I, Cuevas CH, Burgos-Vargas R, et al. Diagnosis of Chronic Gout: Evaluating the American College of Rheumatology Proposal, European League Against Rheumatism Recommendations, and Clinical Judgment. *Journal of Rheumatology*. 2010 Aug;37(8):1743-8. PMID: WOS:000280860600028.

42. Sivera F, Andres M, Carmona L, et al. Multinational evidence-based recommendations for the diagnosis and management of gout: integrating systematic literature review and expert opinion of a broad panel of rheumatologists in the 3e initiative. *Annals of the Rheumatic Diseases*.

2014 Feb;73(2):328-35. PMID:  
WOS:000329488000001.

43. Hasselbacher P. Variation in synovial fluid analysis by hospital laboratories. *Arthritis and Rheumatism*. 1987;30(6):637-42.

44. Park JW, Ko DJ, Yoo JJ, et al. Clinical factors and treatment outcomes associated with failure in the detection of urate crystal in patients with acute gouty arthritis. *Korean Journal of Internal Medicine*. 2014;29(3):361-9.

45. Rigby AS, Wood PH. Serum uric acid levels and gout: what does this herald for the population? *Clin Exp Rheumatol*. 1994 Jul-Aug;12(4):395-400. PMID: 7955603.

46. Segal JB, Albert D. Diagnosis of crystal-induced arthritis by synovial fluid examination for crystals: lessons from an imperfect test. *Arthritis Care Res*. 1999 Dec;12(6):376-80. PMID: 11081008.

47. Chen LX, Schumacher HR. Gout: can we create an evidence-based systematic approach to diagnosis and management? *Best Practice and Research: Clinical Rheumatology*. 2006 August;20(4):673-84. PMID: 2006438744 MEDLINE PMID 16979531 (<http://www.ncbi.nlm.nih.gov/pubmed/16979531>) FULL TEXT LINK <http://dx.doi.org/10.1016/j.berh.2006.05.006>.

48. Hamburger MI, Tesser JRP, Skosey JL, et al. Patterns of gout treatment and related outcomes in us community rheumatology practices: The relation between gout flares, time in treatment, serum uric acid level and urate lowering therapy. *Arthritis and Rheumatism*. 2012 October;64 SUPPL. 10:S808-S9.

49. Dalbeth N. Management of gout in primary care: challenges and potential solutions. *Rheumatology (Oxford)*. 2013 Sep;52(9):1549-50. PMID: 23771950.

50. Harrold LR, Saag KG, Yood RA, et al. Validity of gout diagnoses in administrative data. *Arthritis Care and Research*. 2007 15;57(1):103-8. PMID: 2007089061 MEDLINE PMID 17266097 (<http://www.ncbi.nlm.nih.gov/pubmed/17266097>) FULL TEXT LINK <http://dx.doi.org/10.1002/art.22474>.

## Abbreviations / Acronyms

3e	Evidence, Expertise, Exchange
3ei	Evidence, Expertise, Exchange initiative
ACA	American College of Rheumatology
ACP	American College of Physicians
AHRQ	Agency for Healthcare Research and Quality
ARA	American Rheumatism Association (formerly American College of Rheumatology)
AUC	Area under the curve
CART	Classification and regression tree
CGD	Clinical Gout Disease
CI	Confidence interval
CPPD	Calcium Pyrophosphate Deposition (formerly pseudogout)
CT (Scan)	Computerized tomography
CVD	Cardiovascular disease
DCS	Double-Contour Sign
DECT	Dual-energy computed tomography
EPC	California Evidence-based Practice Center
FN	False negative
FP	False positive
HTN	Hypertension
KQ	Key Question
LDH	Lactate dehydrogenase
LR	Likelihood ratio
LRN	Negative likelihood ratio
LRP	Positive likelihood ratio
MRI	Magnetic resonance imaging
MSU	Monosodium urate
MTP	Metatarsophalangeal
NHANES	National Health and Nutrition Examination Survey
NLR	Nonlikelihood ratio
NPV	Negative predictive value
NSAIDs	Nonsteroidal anti-inflammatory drugs
NY	New York
OA	Osteoarthritis
PCPs	Primary care physicians
PICOTS	Populations, Interventions, Comparators, Outcomes, and Timing
PLM	Polarized light microscopy
PLR	Positive likelihood ratio

PPV	Positive predictive value
RA	Rheumatoid arthritis
ROC	Receiver-operating characteristics
SA	Spondylo-arthropathy
SCEPC	Southern California Evidence-based Practice Center
SFWBC	Synovial Fluid White Blood Cell Count
SpA	Spondyloarthritis
SRs	Systematic Reviews
sUA	Serum uric acid
TEP	Technical Expert Panel
US	Ultrasound
US DHHS	U.S. Department of Health and Human Services
VA	Veterans Administration

